

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 12:53:28 ; Search time 199.974 seconds  
(without alignments)  
10303.209 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_3662\_4146

Perfect score: 485

Sequence: 1 gacgtgtacagcagcgga.....ggagcagctgcacacagga 485

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04: \*  
1: Geneseqn1980s: \*  
2: Geneseqn1990s: \*  
3: Geneseqn2000s: \*  
4: Geneseqn2001as: \*  
5: Geneseqn2001bs: \*  
6: Geneseqn2002s: \*  
7: Geneseqn2003as: \*  
8: Geneseqn2003bs: \*  
9: Geneseqn2003cs: \*  
10: Geneseqn2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	485	100.0	24120	7	ABX11642 Human ser
2	218.6	45.1	1251	9	ADC30509 Human nov
3	218.6	45.1	2488	4	AAH18453 Human CDN
4	200.4	41.3	2768	7	ABX71198 Novel hum
5	74.4	15.3	2737	4	AAI59251 Human pol
6	74.4	15.3	3999	4	AAI61037 Human pol
7	74.4	15.3	5382	9	ADD14722 Human src
8	73.4	15.1	1645	5	AAF24162 Human sec
9	73.4	15.1	1948	2	ABV74346 Human IL-
10	73.4	15.1	1949	2	ABV74347 Human IL-
11	73.4	15.1	2170	6	ABQ54970 Human ova
12	67.8	14.0	724	4	AAH08421 Human CDN
13	67.8	14.0	2184	4	AAH18256 Human CDN
14	65.8	13.6	14061	6	ABV99363 Human NOV
15	65.8	13.6	14109	6	ABV99362 Human NOV
16	64.2	13.2	7564	7	ABZ24581 Human cel
17	63.6	13.1	2259	7	ACC46215 Human dit
18	61.4	12.7	2254	3	AAC93403 Human sec
19	61.4	12.7	2354	5	AAf24183 Human sec
20	61	12.6	716	6	ABK35726 CDNA sequ
21	57.4	11.8	572	5	AA42508 Human CDN
22	54.6	11.3	20565	4	AAK89083 Human dig
23	54.6	11.3	20565	4	AA628908 Human imm

24	54.6	11.3	20565	4	AAI03357	Human rep
25	54.6	11.3	20585	9	ADB31749	Human nov
26	54.2	11.2	1256	6	ABK35725	CDNA sequ
27	53.8	11.1	1005	6	ABN21414	Human ORF
28	53.6	11.1	2884	7	ADA53433	Human cod
29	53.2	11.0	9591	4	AA528906	Human imm
30	53.2	11.0	9591	4	AA528906	Human imm
31	53.2	11.0	9591	9	ADB31747	Human rep
32	53.2	11.0	12415	4	AA528907	Human imm
33	53.2	11.0	12415	4	AAI03356	Human rep
34	53.2	11.0	12415	9	ADB31748	Human nov
35	50.6	10.4	20448	4	ABL19989	Human nov
36	50.6	10.4	24971	4	ABL19988	Drosophil
37	49.2	10.1	81940	4	AA505390	Human tit
38	49.2	10.1	81940	6	ABK64829	Human ben
39	49.2	10.1	93801	8	ABX13540	Human RGS
40	49	10.1	3575	4	AAH57402	Human ske
41	47.2	9.7	3597	4	AAI57881	Human pol
42	47.2	9.7	3603	4	AAI59667	Human pol
43	46	9.5	7328	4	AAK89081	Human dig
44	44.6	9.2	352	5	AAF64470	Novel hum
45	43.4	8.9	11916	4	AAH79279	Streptomy

ALIGNMENTS

RESULT 1  
ABX11642  
ID ABX11642 standard; CDNA; 24120 BP.

XX AC ABX11642;

XX DT 09-MAY-2003 (first entry)

XX DE Human serine/threonine or protein kinase 12599, cDNA.

XX KW Human; ss; gene; serine/threonine kinase; protein kinase; 12599;  
cardiovascular disease; heart failure; myocardial infarction;  
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;  
KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;  
KW haemolytic anaemia; cellular proliferative disorder; cancer;  
KW protein kinase disorder; autoimmune disorder; diabetes mellitus;  
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;  
KW multiple sclerosis.

XX OS Homo sapiens.

XX EH Key Location/Qualifiers

XX FI 5'UTR 1..71

XX FT CDS /\*tag= a

XX FT CDS 72..23978

XX FT CDS /\*tag= b

XX FT 3'UTR /product= "Kinase 12599"

XX FT 3'UTR /note= "This CDS is specifically claimed in claim 2"

XX FT 3'UTR 23979..24120

XX FT 3'UTR /\*tag= c

XX US2002168742-A1.

XX PD 14-NOV-2002.

XX PF 15-FEB-2002; 2002US-00077130.

XX PR 15-FEB-2001; 2001US-0269201P.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Kapeller-Libermann R, Acton SL;

XX DR WPI; 2003-298729/29.

XX DR P-PSDB; ABG76187.

XX XX

PT Novel isolated human protein kinase, designated 59079 or 12599  
PT polypeptide, useful as diagnostic and therapeutic agents for preventing  
PT cardiovascular diseases, proliferative disorders, and protein kinase  
PT disorders.  
XX  
PS Claim 2; Page 58-84; 119pp; English.  
XX  
XX The invention relates to an isolated human serine/threonine or protein  
kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule  
comprising at least 85% identity to the nucleic acids appearing as  
CC ABX11641 and ABX11642 or their complement, a naturally occurring variant  
of the kinases or their fragments. Also included are a non-human host  
cell containing the nucleic acids, an antibody specific for the proteins,  
CC identifying a compound which binds to the kinase (by contacting the  
kinase or a cell expressing the kinase with a test compound and  
determining whether the kinase binds to the test compound) and modulating  
the activity of kinase using the identified compound. The kinases and  
their encoding nucleic acids are useful as diagnostic and therapeutic  
agents for preventing a disease or condition associated with an aberrant  
or unwanted 59079 or 12599 activity in a subject, including  
CC cardiovascular diseases such as heart failure, and myocardial infarction;  
disorders involving blood vessels such as atherosclerosis, and Kaposi's  
CC sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia,  
Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders  
such as cancer; and protein kinase disorders such as autoimmune  
disorders, diabetes mellitus, psoriasis, inflammatory bowel disease,  
rheumatoid arthritis, and multiple sclerosis (many examples of diseases  
and disorders are included in the specification). The kinases, their  
encoding nucleic acids and antibodies are useful in screening assays,  
detection assays (e.g. forensic biology), and predictive medicine (e.g.  
diagnostic assays (e.g. prognostic assays, and monitoring clinical trials and  
pharmacogenomics). The kinases and their encoding nucleic acids are  
useful as query sequences to perform a search against public databases to  
identify other family members or related sequences. The present sequence  
encodes the kinase 12599  
XX  
SQ Sequence 24120 BP; 4719 A; 7265 C; 8278 G; 3858 T; 0 U; 0 Other;  
Query Match 100.0%; Score 485; DB 7; Length 24120;  
Best Local Similarity 100.0%; Pred. No. 6.1e-113;  
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GACGTGGTACAGAGCGGAGAGAGCTGAGCTCAGCTCAAAAGTACGATGGAGTCAA 60  
DB 3662 GACGTGGTACAGAGCGGAGAGAGCTGAGCTCAGCTCAAAAGTACGATGGAGTCAA 3721  
QY 61 GGGCTGCAACAGAGCGTGGTAGTGCAGCAGGTGGGCAAGCAGATGCTGGGGAGTACAG 120  
DB 3722 GGGCTGCAACAGAGCGTGGTAGTGCAGCAGGTGGGCAAGCAGATGCTGGGGAGTACAG 3781  
QY 121 CTGCGAGCTGGGGCCACAGAGTCTCTTTCAACTGCATCATCAGAGCCCAAGGCGGT 180  
DB 3782 CTGCGAGCTGGGGCCACAGAGTCTCTTTCAACTGCATCATCAGAGCCCAAGGCGGT 3841  
QY 181 GTTTCACAGAGCGATGGTGGTCAATAGTGGTGGCAAGCAGTGGGCGGCGATGGCCAC 240  
DB 3842 GTTTCACAGAGCGATGGTGGTCAATAGTGGTGGCAAGCAGTGGGCGGCGATGGCCAC 3901  
QY 241 ACTGAGCTGTGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
DB 3902 ACTGAGCTGTGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3961  
QY 301 GCTGAGCTCCAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAG 360  
DB 3962 GCTGAGCTCCAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAGTCCGAG 4021  
QY 361 GCAGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
DB 4022 GCAGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4081  
QY 421 CTCCTTCACCTGAGTGTTCAGAGCCCAAGCGGCGGTTCGCAAGAGAGAGAGTGGGACAA 480  
DB 4082 CTCCTTCACCTGAGTGTTCAGAGCCCAAGCGGCGGTTCGCAAGAGAGAGAGTGGGACAA 4141

QY 481 CAGGA 485  
DB 4142 CAGGA 4146

## RESULT 2

ADG30509  
ID ADC30509 standard; cDNA; 1251 BP.

XX AC ADC30509;

DT 18-DEC-2003 (first entry)

DE Human novel cDNA sequence, SEQ ID NO:591.

XX Human; diagnostic; drug screening; forensics; gene mapping;  
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
KW ulcers; osteoporosis; autoimmune disease; cancer;  
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;  
KW neuroprotective; antianemic; anticoagulant; thrombolytic; vulnerary;  
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
KW gene therapy; chromosome 17; gene; ss.

OS Homo sapiens.

XX WO2003029271-A2.

PN 10-APR-2003.

PD 24-SEP-2002; 2002WO-US030474.

XX 24-SEP-2001; 2001US-0324631P.

XX (HYSB-) HYSEQ INC.

PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
PI Haley-Vicente D, Drmanac RT;

XX WPI; 2003-371981/35.  
DR P-PSDB; ADC31480.

XX New polynucleotide and polypeptide useful for diagnosing, preventing or  
treating conditions such as neurodegenerative diseases, anemias, platelet  
disorders, wounds, ulcers, osteoporosis, autoimmune diseases or  
cancer.

PS Claim 1; SEQ ID NO 591; 1185pp; English.

XX The invention relates to 971 novel human cDNA sequences (ADC29919-  
ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
invention also relates to nucleic acid sequences over 99% identical with  
the novel human cDNAs. The invention additionally encompasses expression  
vectors and host cells comprising a nucleic acid of the invention; the  
recombinant production of a polypeptide of the invention; an antibody  
against a polypeptide of the invention; a method of detecting  
polynucleotides or polypeptides of the invention; and methods of  
identifying a compound which binds to a polypeptide of the invention. The  
invention further discloses methods of preventing, treating or  
ameliorating a medical condition; kits comprising polynucleotide probes  
and/or monoclonal antibodies for carrying out the methods of the  
invention; methods for the identification of compounds that modulate the  
expression or activity of the polynucleotide and/or polypeptide; and 767  
contig sequences corresponding to the cDNA sequences of the invention  
(ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628  
-ADC33394). The nucleic acids and polypeptides of the invention are  
useful in diagnostics, drug screening, forensics, gene mapping, in the  
identification of mutations responsible for genetic disorders or other  
traits, for assessing biodiversity, and in producing many other types of  
data and products dependent on DNA and amino acid sequences. They are  
also used for treating diseases such as Parkinson's disease, Alzheimer's



CC disease and other neurodegenerative diseases, anaemia, platelet  
CC disorders, wounds, ulcers, osteoporosis, autoimmune diseases or  
CC cancer. The nucleic acids may also be used as hybridisation probes or  
CC primers, and in the recombinant production of a protein. The polypeptides  
CC are also useful in generating antibodies, as molecular weight markers,  
CC and as food supplements. The present sequence represents a specifically  
CC claimed human cDNA sequence of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences.

XX Sequence 1251 BP; 329 A; 293 C; 363 G; 266 T; 0 U; 0 Other;

Query Match 45.1%; Score 218.6; DB 9; Length 1251;  
Best Local Similarity 86.1%; Pred. No. 1.1e-45;  
Matches 242; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 163 CACAGAGCCCAAGGAGTGTGTCACAGGAGCAGTGTGTCATATAGGTGCGGACTGA 222  
Db 891 CCCAGAGCCCAAGGAGTGTGTCACAGGAGCAGCAGCAGGAGGTGCGGCTGA 950  
QY 223 GGCAGGGGCCAGTGCCACACTGAGCTGTGAGGTGGCCAGGCCAGACAGAGGTGAGCTG 282  
Db 951 GCGGGGGCCAGTGCCACAGCTGAGCTGCGAGGTGGCCAGGCCAGACAGAGGTGAGCTG 1010  
QY 283 GTACAAGGATGGGAAGAGCTGAGCTCCAGTTCGAAAGTGGCCATAGAGGTGCGGGCTG 342  
Db 1011 GTACAAGGATGGGAAGAGCTGAGTTCAGCTCGAAAGTGGCGCTGGAGGCGGTGGCTG 1070  
QY 343 CATGGGGCAGCTGTGTTGTCAGGAGGCGGAGGAGGAGTCTGGGAGTACACTGTGA 402  
Db 1071 CACAGGAGGCTGTGTTGTCAGGAGGCGGAGGAGGAGGAGGAGGAGTACACTGTGA 1130  
QY 403 GGCTGGGGGCCAGCGGCTCTCTTCCACCTGGATGTTTCAG 443  
Db 1131 GGCAGGGGTCAGCAGAGCTCTCTTCCGCTGCAGGTGGCAG 1171

## RESULT 3

AAH18453  
ID AAH18453 standard; cDNA; 2488 BP.

XX AAH18453;

XX 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:18549.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or

PT diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 18549; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 2488 BP; 561 A; 637 C; 716 G; 574 T; 0 U; 0 Other;

Query Match 45.1%; Score 218.6; DB 4; Length 2488;

Best Local Similarity 86.1%; Pred. No. 1.3e-45;

Matches 242; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 163 CACAGAGCCCAAGGAGTGTGTCACAGGAGCAGTGTGTCATATAGGTGCGGACTGA 222

Db 891 CCCAGAGCCCAAGTGTGTTGTCACAGGAGCAGCAGCAGGAGGTGCGGCTGA 950

QY 223 GGCAGGGGCCAGTGCCACACTGAGCTGTGAGTGGCCAGGCCAGACAGAGGTGAGCTG 282

Db 951 GCGGGGGCCAGTGCCACAGCTGAGCTGCGAGGTGGCCAGGCCAGACAGAGGTGAGCTG 1010

QY 283 GTACAAGGATGGGAAGAGCTGAGCTCCAGTTCGAAAGTGGCCATAGAGGTGCGGGCTG 342

Db 1011 GTACAAGGATGGGAAGAGCTGAGTTCAGCTCGAAAGTGGCGCTGGAGGCGGTGGCTG 1070

QY 343 CATGGGGCAGCTGTGTTGTCAGGAGGCGGAGGAGGAGTCTGGGAGTACACTGTGA 402

Db 1071 CACAGGAGGCTGTGTTGTCAGGAGGCGGAGGAGGAGGAGTACAGCTGTGA 1130

QY 403 GGCTGGGGGCCAGCGGCTCTCTTCCACCTGGATGTTTCAG 443

Db 1131 GGCAGGGGTCAGCAGAGCTCTCTTCCGCTGCAGGTGGCAG 1171

## RESULT 4

ABX71198

ID ABX71198 standard; cDNA; 2768 BP.

XX AC ABX71198;

XX 05-MAR-2003 (first entry)

XX Novel human cDNA sequence #423.

XX Human; gene; ss; nervous system disorder; peripheral neuropathy;  
KW Huntington's disease; amyotrophic lateral sclerosis; haemophilia;  
KW neurodegenerative disease; Parkinson's disease; Alzheimer's disease;  
KW autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;  
KW insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;  
KW ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;  
KW fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;  
KW coagulation disorder; cancer; tumour; inflammatory disease; septic shock;

KW Crohn's disease; anaphylaxis; proliferation; chemotactic;  
 KW differentiation; stem cell growth factor; haematopoiesis; chemokinetic;  
 KW haemostatic; antiinflammatory; expressed sequence tag; EST.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200281731-A2.  
 XX  
 PD 17-OCT-2002.  
 XX  
 PF 29-JAN-2002; 2002WO-US001222.  
 XX  
 PR 30-JAN-2001; 2001US-00774528.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PA (GOOD/) GOODRICH R W.  
 XX  
 PI Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;  
 XX  
 DR WPI; 2003-058563/05.  
 XX  
 PT Novel polypeptide useful for treating neurodegenerative diseases, myeloid  
 PT or lymphoid cell disorders, bone disorders, mechanical and traumatic  
 PT disorders, coagulation disorders, and inflammatory diseases.  
 XX  
 PS Claim 1; Page; 612pp; English.  
 XX  
 CC This invention relates to the cDNA sequences encoding an isolated novel  
 CC human polypeptide. The protein encoded by the nucleic acid of the  
 CC invention is useful for treating central and peripheral nervous system  
 CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic  
 CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,  
 CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus  
 CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)  
 CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopaenia)  
 CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,  
 CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head  
 CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;  
 CC bacterial, viral or fungal infections; allergic conditions such as  
 CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);  
 CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's  
 CC disease, anaphylaxis). The protein may be used to inhibit the growth,  
 CC infection or function of infectious agents such as bacteria, fungi,  
 CC viruses, or to effect bodily characteristics, biorhythms or circadian  
 CC cycles of rhythms. The protein may also have  
 CC proliferation/differentiation, stem cell growth factor, haematopoiesis  
 CC regulation, immune stimulating or suppressing, chemotactic/chemokinetic,  
 CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory  
 CC activities. The cDNA sequences of the invention are useful for expressing  
 CC recombinant protein for analysis. The present sequence represents a novel  
 CC human cDNA sequence of the invention, this sequence is an expressed  
 CC sequence tag (EST) and was identified using subtractive hybridisation  
 XX  
 SQ Sequence 2768 BP; 456 A; 843 C; 995 G; 468 T; 0 U; 0 Other;  
 XX  
 Query Match 41.3%; Score 200.4; DB 7; Length 2768;  
 Best Local Similarity 69.8%; Pred. No. 5.4e-41;  
 Matches 286; Conservative 0; Mismatches 121; Indels 3; Gaps 1;  
 XX  
 QY 5 TGGTACAGGACGGGAGAACCTCAGCTCCAGCTCAAAAGTACGCATCGAGGTCAGGGC 64  
 DB 2349 TGGTACAGGATGGCATGGAGCTGGGCACCTCCGGTGGAGCGCTTCTTCAGGAGGATGTG 2408  
 QY 65 TGCACACG---AAGCTGTGTAGTCAGACAGTGGGCAAGACAGATGCTGGGGAGTACAGC 121  
 DB 2409 GGGACGGCGGACCGCGCTGGTGGCAGCCACAGTACCAGGCGAGGACCTACTCTCC 2468  
 QY 122 TGGCAGGCTGGGGCCGACAGAGTCTCTTCACTGCACATCAGAGCCCAAGGAGTG 191  
 DB 2469 TCCGCGTGGGCGAGGACTCTGTGGACTTCGGGCTCCGGCTCTGTAGCCCAAGGCGGTG 2528  
 QY 182 TTTCACAGGACGAGTGGTGCATATAGTGGTGGGACTGAGGCGAGGGGGCCAGTGCCACA 241

DB 2529 TTTCACAGGACGAGCGCGCTGCAGGGAGGTGCAGGCTGAGTGGGGCCAGCGCCACG 2588  
 QY 242 CTGAGCTGTGAGGTGGCCCGCCAGCCAGACAGAGGTGACGTGTACAAAGATCGGAAGAG 301  
 DB 2589 CTGAGCTGTGAGGTGGCCCGCCAGGACCATGATGGAGGTGACGTGTACAAAGACGGGAAGAG 2648  
 QY 302 CTGAGCTCCAGTTCCAAAGTGGCATAGAGGCTGGGGGTGTCATCGGCAGCTGTGTGTG 361  
 DB 2649 TTGAGCTCCAGTCCGAAAGTGCAGCTGGAGGCTGTGGGCTGTATCGGAGGCTGTGTG 2708  
 QY 362 CAGCAGGAGGCGCAGCATGCTGGGGAGTACCTGTGAGGCTGGGGG 411  
 DB 2709 CAGCAGGTCGGCCAGCAGACTCCGGATAGTACAGCTGTGAAGCCAGGTG 2758  
 XX  
 RESULT 5  
 AAI59251  
 ID AAI59251 standard; cDNA; 2737 BP.  
 XX  
 AC AAI59251;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 1454.  
 XX  
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US034263.  
 XX  
 PR 23-DEC-1999; 99US-00471275.  
 PR 21-JAN-2000; 2000US-00488725.  
 PR 25-APR-2000; 2000US-00552317.  
 PR 20-JUN-2000; 2000US-00598042.  
 PR 19-JUL-2000; 2000US-00620312.  
 PR 03-AUG-2000; 2000US-00653450.  
 PR 14-SEP-2000; 2000US-00662191.  
 PR 19-OCT-2000; 2000US-00693036.  
 PR 29-NOV-2000; 2000US-00727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
 PI Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI; 2001-442253/47.  
 DR P-PSDB; AAM40095.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders such  
 PT as central nervous system injuries.  
 XX  
 PS Claim 1; SEQ ID NO 1454; 10078pp; English.  
 XX  
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX  
SQ Sequence 2737 BP; 484 A; 840 C; 899 G; 514 T; 0 U; 0 Other;  
  
Query Match 15.3%; Score 74.4; DB 4; Length 2737;  
Best Local Similarity 49.5%; Pred. No. 4.8e-09;  
Matches 220; Conservative 0; Mismatches 221; Indels 3; Gaps 1;  
  
QY 5 TGGTACAGGACGGGAAGCTGAGCTCCAGCTCAAAAGTACGCATGGAGTCAAGGC 64  
Db TGGTACAGGACGGGACGAGGTGAGGAGAGTCACTTCGTGCTGGAGATGAGGG 967  
  
QY 65 TGCACAGAAAGCTGTAGTGCACAGAGTGGGCAAGAGATGTGGGAGTACAGTGC 124  
Db CCCATCGCCGCTGTGCTGCCCGCACCCAGCCCTCAGACGGGGCGAGTTTCACTGC 1027  
  
QY 125 GAGCTGGGGCCAGAGAGTCTCTTCACTGCATCAGAGAGCCAGAGGAGTGT 184  
Db GTGCTGGAGATGAGTGTGCTTCACTGTCACTCAGAGAGTCACTTCCTGTGATC 1087  
  
QY 185 GCCAAGGACGATTTGGTGCAATAGAGTGGGACTGAGGCGAGGGGCCAGTGCACACTG 244  
Db GT---GTATCCAGCGCAAGGTGTATGTGCAGCGGTGCCCTGGAGCGTGTGCTG 1144  
  
QY 245 AGCTGTGAGTGGCCAGGCGCCAGACAGAGTGTGTTACAGAGTGGGAAGACTG 304  
Db ACCTGTGAGTATGCGGCCCTTGGCAGAGTGTGCGTGGACCAAGGATGGAGAGAGTG 1204  
  
QY 305 AGCTCCAGTTCGAAAGTCCGATAGAGCTCGGGCTCATGCGGAGCTGTGTGTGAG 364  
Db GTGAGAGCCCGCGCTGCTCTGCAGAGAGGACACTGTCCGCGCGCTGTGTGCTGCC 1264  
  
QY 365 CAGGACGCGCAGGAGATGTGTTGGGAGTACACCTGTGAGGCTGGGGCCAGCGCTCTCC 424  
Db GCTGTCCAGCTCGAGGACTCCGGCGAGTACTTGTGTGAATTCAGATGAGTCGGCTCC 1324  
  
QY 425 TTCCACCTGGATTTTCAGAGCC 448  
Db TTCACTGTACCGTCACAGAACCC 1348  
  
RESULT 6  
AA161037/c  
ID AA161037 standard; cDNA; 3999 BP.  
XX  
AC AA161037;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 5026.  
XX  
KW Human; nototropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US034263.  
XX  
PR 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
PR 20-JUN-2000; 2000US-00538042.  
PR 19-JUL-2000; 2000US-00620312.  
PR 03-AUG-2000; 2000US-00653450.  
PR 14-SEP-2000; 2000US-00662191.  
PR 19-OCT-2000; 2000US-00693036.  
PR 29-NOV-2000; 2000US-00727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
XX Zhou J, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
XX Zhou P, Godrich R, Drmanac RT;  
XX  
XX WPI; 2001-442253/47.  
XX  
XX P-PSDB; AA41881.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders such  
XX as central nervous system injuries.  
XX  
XX Claim 1; SEQ ID NO 5026; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AA157798-AA161369) and the  
XX encoded polypeptides (AA38642-AA42213) with nototropic, and the  
XX immunosuppressant and cytostatic activity. The polynucleotides are useful  
XX in gene therapy. A composition containing a polypeptide or polynucleotide  
XX of the invention may be used to treat diseases of the peripheral nervous  
XX system, such as peripheral nervous injuries, peripheral neuropathy and  
XX localised neuropathies and central nervous system diseases, such as  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
XX utilisation of the activities such as: Immune system suppression,  
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
XX assays for receptor activity, arthritis and inflammation, leukaemias and  
XX C.N.S disorders. Note: The sequence data for this patent did not form  
XX part of the printed specification  
XX  
SQ Sequence 3999 BP; 726 A; 1332 C; 1238 G; 703 T; 0 U; 0 Other;  
  
Query Match 15.3%; Score 74.4; DB 4; Length 3999;  
Best Local Similarity 49.5%; Pred. No. 5.3e-09;  
Matches 220; Conservative 0; Mismatches 221; Indels 3; Gaps 1;  
  
QY 5 TGGTACAGGACGGGAGAGCTGAGCTCCAGCTCAAAAGTACGCATGGAGTCAAGGC 64  
Db TGGTACAGGACGGGACGAGGTGAGGAGAGTCACTTCGTGCTGTGAGATGAGGG 3309  
  
QY 65 TGCACACGAGGCTGTGTAGTGCAGAGGTGGGCAAGCAGATGCTGGGAGTACAGCTGC 124  
Db CCCATCGCCGCTGTGCTGCCCGCACCCAGCCCTCAGACGGGGCGAGTTTCACTGC 3249  
  
QY 125 GAGGCTGGGGCCAGAGAGTCTCTTTCACTGCATCAGAGAGCCCAAGGAGTGT 184  
Db GTGCTGGAGATGAGTGTGCTTCACTGTCACTGTCACTGTCACTGTCTCTCTGTGATC 3189  
  
QY 185 GCCAAGGACGATTTGGTGCAATAGAGTGGGCGAGTGGAGGAGGAGGAGGAGG 244  
Db GT---GTATCCAGCGCAAGGTGTATGTGGAGCGCTGCGCTGTGAGCGTGTGCTG 3132  
  
QY 245 AGCTGTGAGTGGCCAGGCGCCAGACAGAGGTGACCGTGTGTACAGGATGGGAAGAGCTG 304  
Db ACCTGTGAGTATGCGGCCCTTGGCAGAGTGTGCTGGACCAAGGATGAGAGGAGTG 3072  
  
QY 305 AGCTCCAGTTCGAAAGTGGCATAGAGGCTGCCGGCTCATGCGGAGCTGTGTGTG 364  
Db GTGAGAGCCCGCGCTGCTCTGCAGAGAGGACACTGTCCGCGCGCTGTGTGCTGCC 3012  
  
QY 365 CAGGACGCGCAGGAGATGTGGGAGTACACCTGTGAGGCTGGGGCCAGCGCTCTCC 424  
Db GCTGTCCAGCTCGAGGACTCCGGCGAGTACTTGTGTGAATTCAGATGAGTCGGCTCC 2952

Qy 425 TTCCACCTGGATGTTTCAGAGCCC 448  
||| | ||| ||| |||  
Db 2951 TTCACTGTCAACCGTCACAGAACC 2928

RESULT 7  
ADD14722

ID ADD14722 standard: cDNA: 5382 bp.

AC ADD14722:

01-JAN-2004 (first entry)

Human src biomarker no] vnucl] ectide seo ID NO.115

XX predictor set; protein tyrosine kinase activity modulator;  
KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;  
KW gene therapy; drug sensitivity; genetic profile; cancer; human; gene; ss.

xx  
OS  
Homo sapiens.

XX PN W02003062395-A2

31-III-2003

XX  
DE 17-TAN-2003: 2003W0-IIS001991

XX  
19 TAN 2002. 2002020 03F00015

XX

XX

DR P-PSDB; ADD14123.

PT New polynucleotides and polypeptides for predicting the activity of  
PT compounds that interact with protein tyrosine kinases and/or protein  
PT tyrosine kinase pathways.

PS Claim 2: SEQ ID NO 116: 139pp: English: XX

The present invention describes a predictor set comprising a plurality of polynucleotides or polypeptides whose expression pattern is predictive of the response of cells to treatment with a compound that modulates protein tyrosine kinase activity or members of the protein tyrosine kinase pathway. Also described: (1) predicting whether a compound is capable of modulating the activity of cells, comprising obtaining a sample of cells, determining whether the cells express a plurality of markers, and correlating the expression of the markers to the compound's ability to modulate the activity of the cells; (2) a plurality of cell lines for identifying polynucleotides and polypeptides whose expression levels correlate with compound sensitivity or resistance of cells associated with a disease state; and (3) identifying polynucleotides and polypeptides that predict compound sensitivity or resistance of cells associated with a disease state, comprising subjecting the plurality of cell lines to one or more compounds, analysing the expression pattern of a microarray of polynucleotides or polypeptides, and selecting polynucleotides or polypeptides that predict the sensitivity or resistance of cells associated with a disease state by using the expression pattern of the microarray. The polynucleotides and polypeptides have cytostatic activities, and can be used in gene therapy. The polynucleotides and polypeptides are useful in predicting the activity of compounds that interact with protein tyrosine kinases and/or protein tyrosine kinase pathways. These may be used in determining drug sensitivity in patients to allow the development of individualized genetic profiles which aid in treating diseases and disorders (e.g. cancer) based on patient response at a molecular level. The present sequence is used in the exemplification of the present invention.

Sequence 5382 BP; 943 A; 1699 C; 1798 G; 942 T; 0 U; 0 Other;  
SQ

### Query Match

15.3%; Score 74.4; DB 9; Length 5382;

[illegible]

RESULT 8  
AAF24162

AAF24162  
ID AAF24162 standard: DNA: 1645 bp.

XX  
AC  
DAF24162.

XX  
DT 23-MAR-2001 (first sent via)

XX Human generated protein DN2 #12

XX  
KW Secreted protein; gene therapy; vaccine; cancer; leukemia;  
KW autoimmune disease; allergy; inflammation; graft rejection;  
KW hyperproliferation; cardiovascular; infection; ds.

XX  
QS  
Homo sapiens.

XX  
DN WO30007E37E-21[illegible][illegible]

XXXX

[illegible]

(HUMAN) HUMAN GENOME SCI INC.  
 FAX  
 XX

PI	Ruben SM, Birse CE, Duan RD, Soppet DR, Rosen CA, Shi Y;
PI	Lafleur DW, Olsen HS, Ebner R, Florence KA, Ni J, Young PE;

XX  
WPT: 2001-061741/07

PT Nucleic acids encoding 26 human secreted polypeptides, useful e.g. for preventing, diagnosing and/or treating cancers and for promoting wound healing.

xx  
ps  
Claim 1: Page 425-426: 530pp: English:

CC The present invention relates to 26 secreted human proteins. The proteins  
 CC may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate polypeptide expression. For example, they  
 CC may be used in gene therapy or in vaccines. Typical of diseases which are  
 CC potentially treatable are cancers (including leukemia), autoimmune  
 CC diseases, allergies, inflammation, graft rejection, hyperproliferation,  
 CC cardiovascular diseases (particularly critical limb ischemia and coronary  
 CC disease) and any involving abnormal angiogenesis, neurodegeneration  
 CC and/or infectious diseases  
 XX  
 XX Sequence 1645 BP; 338 A; 474 C; 506 G; 327 T; 0 U; 0 Other;  
 XX  
 XX Query Match 15.1%; Score 73.4; DB 5; Length 1645;  
 XX Best Local Similarity 49.4%; Pred. No. 7.7e-09;  
 XX Matches 219; Conservative 0; Mismatches 221; Indels 3; Gaps 1;  
 XX  
 QY 5 TGGTACAGAGCGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGCATGGAGGTCAAGGC 64  
 Db 905 TGGTACAGAGCGGCGAGAGGTGGAGAGAGTCACTTCGTGTCTGGAGATGAGGG 964  
 QY 55 TGCACAGAGGCTGTAGTCAGCAGGTGGGCAAGCAGATGCTGGGAGTACAGCTGC 124  
 Db 965 CCCATCCGCCCTGTGTCTGCCGCCACCCATCCCTCAGACGGGGCGAGTTTCAGTGC 1024  
 QY 125 GAGGCTGGGGCCAGAGAGTCTCTTTCACTGCACATCAGAGCCCCAAGGCGAGTGT 184  
 Db 1025 GTCCGTGAGATGAGTGTGCCTACTTCACTGTCCACATCAGAGCTCTCTCTGTGATC 1084  
 QY 185 GCCAAGAGAGCTGTGTGTGATAGTGTGGGACTGAGGAGGGCCAGTCCACACTG 244  
 Db 1085 GT--GTATCCAGCGGCAAGGTGTATGTGCAGCCGTGCGCTGGAGCGTGTGTCTG 1141  
 QY 245 AGCTGTAGGTGGCCAGGCCAGACAGAGTGTGCTGTACAAAGGATGGAAGAAGCTG 304  
 Db 1142 ACCTGTAGCTATGCGGCCCTGGCAGAGTGTGCTGTGACCAAGGATGGAGAGGTG 1201  
 QY 305 AGCTCCAGTTGAAAGTCCGATAGAGCTGCGGCTGCATGCGGAGCTGTGTGTGAG 364  
 Db 1202 GTGGAGAGCCCGCGCTGCTCTGTGCAAGGAAGACACTGTCCCGCCGCTGTGTGCC 1261  
 QY 365 CAGGAGCGCCAGGACAGTGTGGGAGTACACCTGTGAGGTGGGGCCAGCGCTTCC 424  
 Db 1262 GCTGTCCAGCTCGAGGATCCGGCGAGTACTTGTGTGAAATGACGATGAGTCGCGCTCC 1321  
 QY 425 TTCCACCTGGATTTTCAGAGCC 447  
 Db 1322 TTCACTGTACCGTTCACAGAGTC 1344  
 XX  
 XX RESULT 9  
 XX ID ABV74346  
 XX ABV74346 standard; cDNA; 1948 BP.  
 XX AC ABV74346;  
 XX XX  
 XX DT 21-FEB-2003 (first entry)  
 XX DE Human IL-1RRP3 encoding cDNA SEQ ID NO 1.  
 XX KW Human; interleukin 1 receptor related protein 3; IL-1RRP3; inflammation;  
 XX arthritis; septicemia; autoimmune disease; transplant rejection;  
 XX infection; stroke; AIDS; osteoporosis; cancer; Alzheimer's disease;  
 XX anti-inflammatory; antiarthritic; antibacterial; immunosuppressive;  
 XX cerebroprotective; vasotropic; antiviral; osteopathic; cytostatic;  
 XX antiarteriosclerotic; neurotropic; neuroprotective; chromosome 2q33; gene;  
 XX ss.  
 XX OS Homo sapiens.  
 XX XX  
 XX FH Key Location/Qualifiers  
 XX CDS 413..1699  
 XX FT /\*tag= a  
 XX FT /product= "IL-1RRP3"  
 XX FT

XX WO9918990-A1.  
 PN 22-AFR-1999.  
 XX 09-OCT-1998; 98WO-US021347.  
 PF 10-OCT-1997; 97US-0061720P.  
 XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA Young PR;  
 XX WPI; 1999-277442/23.  
 DR P-PSDB; ABB98343.  
 XX Novel interleukin 1 receptor related protein 3 (IL-1RRP3) or variants  
 FT for, e.g. treatment of Alzheimer's disease.  
 XX Claim 2; Page 30-31; 43pp; English.  
 XX The invention relates to polynucleotides encoding interleukin 1 receptor  
 CC related protein 3 (IL-1RRP3) or variants. The polynucleotides are used to  
 CC diagnose a disease or susceptibility to a disease in a subject related to  
 CC expression or activity of the IL-1RRP3 to screen for compounds which  
 CC stimulate or inhibit the function of IL-1RRP3 or to treat a subject in  
 CC need of enhanced or inhibited IL1RRP3 activity. Specifically mentioned in  
 CC the specification is the use of the IL-3RRP3 polynucleotide or  
 CC polypeptide for treatment of chronic and acute inflammation, arthritis,  
 CC septicemia, autoimmune disease, transplant rejection, graft-versus-host  
 CC disease, infection, stroke, ischaemia, acute respiratory disease  
 CC syndrome, restenosis, brain injury, AIDS, bone diseases e.g.  
 CC osteoporosis, cancer, atherosclerosis and Alzheimer's disease. The  
 CC present sequence is that of a human IL-1RRP3 encoding cDNA of the  
 CC invention. The IL-1RRP3 gene is located on chromosome 2q33  
 XX  
 XX Sequence 1948 BP; 397 A; 561 C; 600 G; 386 T; 0 U; 4 Other;  
 XX  
 XX Query Match 15.1%; Score 73.4; DB 2; Length 1948;  
 XX Best Local Similarity 49.4%; Pred. No. 8e-09;  
 XX Matches 219; Conservative 0; Mismatches 221; Indels 3; Gaps 1;  
 XX  
 QY 5 TGGTACAGAGCGGGAAGAGTGTGAGCTCCAGCTCAAAAGTACGCATGGAGGTCAAGGC 64  
 Db 1139 TGGTACAGAGCGGCGAGAGGTGGAGGAGTGTGCTGTGTGTGAGATGAGGG 1198  
 QY 65 TGCACACGAGGCTGTGTGTGAGTGTGAGCAGAGTGGGCAAGCAGATGCTGGGAGTACAGCTGC 124  
 Db 1199 CCCATCGCGCGCTGTGTGTGCTGCCCGCCAGCCCTCAGACGGGGCGAGTTTCAGTGC 1258  
 QY 125 GAGGCTGGGGCCAGAGAGTCTCTTTCACTGCACATCAGAGAGCCCAAGGAGTGT 184  
 Db 1259 GTCCGTGGAGATGAGTGTGCTTACTTCACTGTCAACATCAGACAGCTCTCCTCGTGATC 1318  
 QY 185 GCCAAGGAGCAGTTGTGTGCATATGAGTGTGCGGACTGAGGCGAGGGCCAGTGCACACTG 244  
 Db 1319 GT---GTATCCAGCGCAGAGTGTATGTGGCAGCGCTGCGCTGGAGCGTGTGTGCTG 1375  
 QY 245 AGCTGTGAGTGGCCAGAGCCAGAGAGTGTACGTGTACAGAGTGGAGAGAGCTG 304  
 Db 1376 ACCTGTGAGTATATCCGGCCCTCGGAGAGGTGCGTGGACCAAGAGATGAGAGGAGGTG 1435  
 QY 305 AGCTCCAGTTTCAAAAGTGGCCATAGAGGCTGCGGGGTGCATGCGGAGCTGGTGTGTGAG 364  
 Db 1436 GTGGAGAGCCCGCGCTGTCTCTGCGAAGAGAGACACTCTCCGCCCGCTGGTGTGCTCC 1495  
 QY 365 CAGCAGGCCAGGAGATGTGTGGGAGTACACCTGTGAGGCTGGGGCCAGCGGCTCTCC 424  
 Db 1496 GCTGTCCAGCTCGAGGACTCCGGCGAGTACTTGTGTGAAATGACGATGATGCGGCTCC 1555  
 QY 425 TTCCACCTGGATGTTTCAGAGCC 447  
 Db 1556 TTCACCTGCCGTCACAGAGTC 1578

RESULT 10  
 ABV74347  
 ID ABV74347 standard; cDNA; 1949 BP.  
 XX AC ABV74347;  
 XX DT 21-FEB-2003 (first entry)  
 XX DE Human IL-1RRP3 encoding cDNA SEQ ID NO 3.  
 XX KW Human; interleukin 1 receptor related protein 3; IL-1RRP3; inflammation;  
 KW arthritis; septicemia; autoimmune disease; transplant rejection;  
 KW infection; stroke; AIDS; osteoporosis; cancer; Alzheimer's disease;  
 KW anti-inflammatory; antiarthritic; antibacterial; immunosuppressive;  
 KW cerebroprotective; vasotropic; antiviral; osteopathic; cytostatic;  
 KW antiarteriosclerotic; nootropic; neuroprotective; chromosome 2q33; gene;  
 KW ss.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 FT CDS 414..1700  
 FT /\*tag= a  
 FT /product= "IL-1RRP3"  
 XX PN WC9918990-A1.  
 XX PD 22-APR-1999.  
 XX PF 09-OCT-1998; 98WO-US021347.  
 XX PR 10-OCT-1997; 97US-0061720P.  
 XX PA (SMK ) SMITHKLINE BEECHAM CORP.  
 XX PI Young PR;  
 XX WPI; 1999-277442/23.  
 DR P-PSDB; ABB98344.  
 XX Novel interleukin 1 receptor related protein 3 (IL-1RRP3) or variants  
 PT for, e.g. treatment of Alzheimer's disease.  
 XX Claim 13; Page 31-32; 43pp; English.  
 CC The invention relates to polynucleotides encoding interleukin 1 receptor  
 CC related protein 3 (IL-1RRP3) or variants. The polynucleotides are used to  
 CC diagnose a disease or susceptibility to a disease in a subject related  
 CC expression or activity of the IL-1RRP3 to screen for compounds which  
 CC stimulate or inhibit the function of IL-1RRP3 or to treat a subject in  
 CC need of enhanced or inhibited IL1RRP3 activity. Specifically mentioned in  
 CC the specification is the use of the IL-3RRP3 polynucleotide or  
 CC polypeptide for treatment of chronic and acute inflammation, arthritis,  
 CC septicemia, autoimmune disease, transplant rejection, graft-versus-host  
 CC disease, infection, stroke, ischaemia, acute respiratory disease  
 CC syndrome, restenosis, brain injury, AIDS, bone diseases e.g.  
 CC osteoporosis, cancer, atherosclerosis and Alzheimer's disease. The  
 CC present sequence is that of a human IL-1RRP3 encoding cDNA of the  
 CC invention. The IL-1RRP3 gene is located on chromosome 2q33  
 XX  
 SQ Sequence 1949 BP; 398 A; 561 C; 600 G; 386 T; 0 U; 4 Other;  
 Query Match 15.1%; Score 73.4; DB 2; Length 1949;  
 Best Local Similarity 49.4%; Pred. No. 8e-09;  
 Matches 219; Conservative 0; Mismatches 221; Indels 3; Gaps 1;  
 QY 5 TGGTACAAAGGACGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGGAGGTCAAGGCG 64  
 DB 1140 TGGTACAAAGGACGGGAGGAGGTGGAGAGAGTGACTTCGTGTGCTGGAGATGAGGGG 1199  
 QY 65 TGCACACGAGGCTGGTGTAGTGCAGCAGGTGGCAAGACAGATGCTGGGAGTACAGCTCC 124

Db 1200 CCCATCGCGCTGGTGTGCGCGCCACCCAGCCCTCAGACGGGGGAGTTTCAGTGC 1259  
 QY 125 GAGGCTGGGGGCCAGAGAGTCTCTTTCAACTGCACATCACAGAGCCCAAGGCAGTGTGTT 184  
 Db 1260 GTGCTGGAGATGAGTGTGCTTACTTCACTGTCAACATCACAGACGTCTCTCTCGTGGATC 1319  
 QY 185 GCCAAGGACGATTGTCATTAATGAGGTGCGGACTGAGGCGAGGGGCCAGTCCACACTG 244  
 Db 1320 GT---GTATCCAGCGGCAAGGTGTATGTGGCAGCGCTGCGCTGTGAGCGTGTGTGCTG 1376  
 QY 245 AGCTGTGAGGTGGCCCGCCAGGCCAGACAGAGGTGACGTGTACAAAGATGGGAAGAGCTG 304  
 Db 1377 ACTGTGAGCTATGCGGCGCTTGGCAGAGGTGCGCTGGACCAAGATGGAGAGAGGTG 1436  
 QY 305 AGCTCAGTTCGAAAGTGGGCATAGAGGTGCGGGCTGCATGCGGCAGCTGGTGTGTGAG 364  
 Db 1437 GTGGAGAGCCCGCGCTGCTCTCTGCAGAGGAGACACTGTCCGCGCTGTGTGTGCTGCC 1496  
 QY 365 CAGGACAGGCCAGGACAGATGCTGGGAGTACACTGTGAGGCTGGGGGCCAGGGCTCTCC 424  
 Db 1497 GCTGTCCAGCTCGAGGACTCCGCGAGTACTTGTGTGAATTGACATGATCGCGCTCC 1556  
 QY 425 TTCCACCTGGATTTTCAGAGCC 447  
 Db 1557 TTCACTGTCAACGTCACAGAGTC 1579

RESULT 11  
 ABQ54970  
 ID ABQ54970 standard; cDNA; 2170 BP.  
 XX AC ABQ54970;  
 DT 22-AUG-2002 (first entry)  
 XX DE Human ovarian antigen HNOCH54 cDNA, SEQ ID NO:850.  
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
 KW inflammatory condition; immune disorder; blood disorder;  
 KW cardiovascular disorder; respiratory disorder; neurological disorder;  
 KW gastrointestinal disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antiinflammatory; gynaecological; reproductive; gene; ss.  
 XX OS Homo sapiens.  
 XX WO200200677-A1.  
 XX 03-JAN-2002.  
 XX 07-JUN-2001; 2001WO-US018569.  
 XX 07-JUN-2000; 2000US-0209467P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Birse CE, Rosen CA;  
 XX WPI: 2002-147878/19.  
 XX P-PSDB; ABP41893.  
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
 PT cancer), immune disorders, cardiovascular disorders and neurological  
 PT diseases.  
 PS Claim 1; SEQ ID NO 850; 2922pp; English.

CC The invention relates to 2175 novel human ovarian antigens (ABP41054-  
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
CC to the sequences of the invention. The invention additionally relates to  
CC recombinant vectors and host cells comprising human ovarian antigen  
CC polynucleotides, antibodies against human ovarian antigens, and the use  
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
CC treating, prognosing or preventing various ovary and/or breast-related  
CC disorders. Such conditions include ovarian cancer and breast cancer, and  
CC metastatic tumours of ovarian or breast origin, reproductive system  
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis), and toxic  
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
CC vaginitis), immune disorders (e.g., congenital and acquired  
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
CC respiratory disorders, neurological disorders, gastrointestinal disorders  
CC and urinary system disorders. Ovarian antigen polypeptides and  
CC polynucleotides may also be used in screening for compounds which  
CC modulate ovarian antigen expression or activity. The polynucleotides may  
CC further be used for gene therapy, chromosome mapping, in the  
CC identification of individuals and in forensic analysis, and the  
CC polypeptides may be used as food additives or to prepare antibodies  
CC useful in disease diagnosis, drug targeting and phenotyping. The present  
CC sequence represents cDNA encoding a human ovarian antigen of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pat\_sequences  
CC XX

XX Sequence 2170 BP; 456 A; 624 C; 679 G; 411 T; 0 U; 0 Other;

Query Match 15.1%; Score 73.4; DB 6; Length 2170;  
Best Local Similarity 49.4%; Pred. No. 8.2e-09;  
Matches 219; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 5 TGGTACAGAGCGGAGAAAGCTGAGCTTCAGCTCAAGTACCAAGTACGAGGTCAAGGGC 64  
DB |||||  
1396 TGGTACAGAGCGGAGGAGGTGGAGAGTGAATTCGTGCTGGAGAAATGAGGGG 1455  
QY 65 TGCACACGAAGCTGTGAGTACGAGGTGGGCAACGATGCTGGGAGTACAGCTGC 124  
DB |||||  
1456 CCCATCCGCCCGCTGCTGCCCGCCACCCAGCCCTCAGACGGGGGAGTTTCAGTGC 1515  
QY 125 GAGGTGGGGGCCAGAGAGTCTCTTTCAACTGCACATCAGAGCCCAAGGCGAGTGT 184  
DB |||||  
1516 GTGCTGGAGATGAGTGTGCTACTTCACTGTCCACATCAGACGCTCTCTCTGGATC 1575  
QY 185 GCCAAGAGCAGTGTGTCATATGAGTGGGACTGAGGCGGCGCCAGTCCACACTG 244  
DB |||||  
1576 GT---GTATCCAGCGGCAAGGTGTATGTGGCAGCCGTGCCCTGGAGCGTGTGCTG 1632  
QY 245 AGCTGTGAGTGGGCCAGGCCAGACAGAGTGTGCTGTGTTACAGGATGGGAAGAGT 304  
DB |||||  
1633 ACCTGTGAGTATGCGGCCCTGGCGAGAGTGTGCTGTGACCAAGGATGGAGAGGTG 1692  
QY 305 AGCTCCAGTTGAAAGTGGCATAGAGCTGCGGCGTGCATGCGGCGAGCTGTGTGTCAG 364  
DB |||||  
1693 GTGGAGAGCCCGCGCTGCTCTCTGTCAGAGGAAGACACTGTCCCGCGCTGTGTGTCGCC 1752  
QY 365 GAGGAGAGCCAGCAGATGCTGGGAGTACACCTGTGAGGCTGGGGCCAGCGGCTTCC 424  
DB |||||  
1753 GTGTGTCAGCTCGAGACTCCCGGCGAGTACTTGTGTGAATTGACATGAGTCGCGCTCC 1812  
QY 425 TTCCACCTGGATGTTTCAGAGCC 447  
DB |||||  
1813 TTCACGTGTCACGTCACAGAGTC 1835

RESULT 12  
AAH08421  
ID AAH08421 standard; cDNA; 724 BP.  
XX

AC AAH08421;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA clone (5'-primer) SEQ ID NO:5256.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 93JP-00248036.  
PR 27-AUG-1999; 95JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
XX length cDNAs defined in the specification, and for the detection and/or  
XX diagnosis of the abnormality of the proteins encoded by the full-length  
XX cDNAs.  
XX  
XX Claim 1; SEQ ID NO 5256; 2537pp + Sequence Listing; English.  
XX  
XX The present invention describes primer sets for synthesising 5602 full-  
XX length cDNAs defined in the specification. Where a primer set comprises:  
XX (a) an oligo-dr primer and an oligonucleotide complementary to the  
XX complementary strand of a polynucleotide which comprises one of the 5602  
XX nucleotide sequences defined in the specification, where the  
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX of an oligonucleotide comprising a sequence complementary to the  
XX complementary strand of a polynucleotide which comprises a 5'-end  
XX sequence and an oligonucleotide comprising a sequence complementary to a  
XX polynucleotide which comprises a 3'-end sequence, where the  
XX oligonucleotide comprises at least 15 nucleotides and the combination of  
XX the 5'-end sequence/3'-end sequence is selected from those defined in the  
XX specification. The primer sets can be used in antisense therapy and in  
XX gene therapy. The primers are useful for synthesising polynucleotides,  
XX particularly full-length cDNAs. The primers are also useful for the  
XX detection and/or diagnosis of the abnormality of the proteins encoded by  
XX the full-length cDNAs. The primers allow obtaining of the full-length  
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893  
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent  
XX oligonucleotides, all of which are used in the exemplification of the  
XX present invention  
XX  
SQ Sequence 724 BP; 116 A; 218 C; 253 G; 134 T; 0 U; 3 Other;

Query Match 14.0%; Score 67.8; DB 4; Length 724;  
Best Local Similarity 46.8%; Pred. No. 1.7e-07;  
Matches 213; Conservative 0; Mismatches 242; Indels 0; Gaps 0;

QY 5 TGGTACAGAGCGGAGAGCTGAGCTCCAGCTCAAGTACCAAGTACGATCGTGGGAGTACAGTGC 64  
DB |||||  
45 TGGTACAGAGTGGCTTGGAAAGTGGAGGAGGAGCGCCCTGGTGTCTGAGAGGATGGG 104  
QY 65 TGCACACGAAGCTGTGAGTGCAGAGGTGGGCAAGAGCAGATCGTGGGAGTACAGTGC 124  
DB |||||  
105 CCAGCTGCCGCTGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 164









Db 360 CCTGAAGACGCCGCTGTACCTGCCACGTGGGCTCGAGGAGACCCGGGCCCGGGTC 419

Qy 434 GATTTTCAGAGCCCAAGCGGTGTTTGCAAGAGCAGCTGSCA 478

Db 420 CGCGTGACGATCTGTCACGTGGGCATCACCAAGGGGCTGAAGACA 464

Search completed: September 19, 2004, 15:42:36  
Job time : 202.974 secs

**This Page Blank (uspto)**





FILE REFERENCE: CL000927-CIP-DIV  
; CURRENT APPLICATION NUMBER: US/10/274,978  
; CURRENT FILING DATE: 2002-10-22  
; PRIOR APPLICATION NUMBER: 09/858,664  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 09/711,134  
; PRIOR FILING DATE: 2000-11-14  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 5207  
; TYPE: DNA  
; ORGANISM: Human  
US-10-274-978-1

Query Match  
Best Local Similarity 57.4%; Pred. No. 0.13;  
Matches 74; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1 GACGTGGTACAGGACGGGAGAGCTGAGCTCCAGCTCAAAAGTACGCATGGAGGTCAA 60  
DB 363 GACCTGGTACAGGACAGCTGCTGAGCAGACCCGGCTTAGCCAGCAGCAAGA 422

QY 61 GGGCTGCACAGGAGCTGGTGTAGTCAGCAGGTGGGCAAGACATGCTGGGGAGTACAG 120  
DB 423 AGGCACCATACTCCCTGGTGTGAGGCATGTGGCTCGAAGGATGCCGGCGTTTACAC 482

QY 121 CTGCGAGGC 129  
DB 483 CTGCTGGC 491

RESULT 5  
US-10-274-978-3  
; Sequence 3, Application US/10274978  
; Patent No. 6670164  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui, et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; MOLECULE TYPE: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL000927-CIP-DIV  
; CURRENT APPLICATION NUMBER: US/10/274,978  
; CURRENT FILING DATE: 2002-10-22  
; PRIOR APPLICATION NUMBER: 09/858,664  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 09/711,134  
; PRIOR FILING DATE: 2000-11-14  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 5207  
; TYPE: DNA  
; ORGANISM: Human  
US-10-274-978-3

Query Match  
Best Local Similarity 57.4%; Pred. No. 0.13;  
Matches 74; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1 GACGTGGTACAGGACGGGAGAGCTGAGCTCCAGCTCAAAAGTACGCATGGAGGTCAA 60  
DB 363 GACCTGGTACAGGACAGCTGCTGAGCAGACCCGGCTTAGCCAGCAGCAAGA 422

QY 61 GGGCTGCACAGGAGCTGGTGTAGTCAGCAGGTGGGCAAGACATGCTGGGGAGTACAG 120  
DB 423 AGGCACCATACTCCCTGGTGTGAGGCATGTGGCTCGAAGGATGCCGGCGTTTACAC 482

QY 121 CTGCGAGGC 129  
DB 483 CTGCTGGC 491

RESULT 6  
US-08-209-747-1  
; Sequence 1, Application US/08209747  
; Patent No. 5733771  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; APPLICANT: Celgin, Mark  
; TITLE OF INVENTION: CDNAS Encoding Minor Ampullate Spider  
; TITLE OF INVENTION: Silk Proteins  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-3487  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/209,747  
; FILING DATE: 14-MAR-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 1447-104P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-205-8000  
; TELEFAX: 703-205-8050  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2793 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Nephila clavipes  
; TISSUE TYPE: minor ampullate gland  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 183..2675  
; OTHER INFORMATION: /product= "N. clavipes minor  
; OTHER INFORMATION: ampullate silk protein"  
US-08-209-747-1

Query Match  
Best Local Similarity 45.8%; Pred. No. 0.15;  
Matches 140; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 167 GAGCCCAAGCAGCTGTTGCCAAGGAGCAGTTGGTGCATAATGAGGTGGGACTGAGGCA 226  
DB 520 GTCCGGAGCCGAGAGCTGGTGTCTGCAGCTCTGTGAGCAGGAGCCGGAGGCGCTGGTG 579

QY 227 GGGGCCACAGTCCCACTGAGCTGTGAGGTGGCCAGCCAGACAGAGGTGACGTGGTAC 286  
DB 580 GTTACGCTAGAGTGTGGTGTGGAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 639

QY 287 AAGATCGGAGAGTGTGAGTCCAGTCCAGTTCGAAGTGGCATAGAGCTCGGGCTCATG 346  
DB 640 GTGGTGTGATATGTGGGCAAGGCGGATATGTGTCGGAGCAGGAGTGTGGTGTGGTGTGG 699

QY 347 CGGCAGCTGTGTGTGAGCAGGAGGAGGAGGAGTGTGGGAGTACACCTGTGAGGCT 406  
DB 700 CTGCTGTGTGTGAGGAGGAGGAGGAGTGTGGGAGTACGCTAGAGTGTGTGTGTGGAG 759

QY 407 GGGGGCCAGCGGCTCTCTTCCACCTGGATGTTTACAGCCCAAGGGGTGTTTCAAG 466

Db 760 CAGGAGCGCTGGGGTGTGGAGCTGGAGGCTACGGTGGTCAAGGTGGGTACGGTGCCG 819

QY 467 GAGCAG 472

Db 820 GAGCAG 825

RESULT 7

US-08-458-298-1

; Sequence 1, Application US/08458298

; Patent No. 5756677

; GENERAL INFORMATION:

; APPLICANT: Lewis, Randolph V.

; APPLICANT: Colgin, Mark

; TITLE OF INVENTION: CDNAs Encoding Minor Ampullate Spider

; TITLE OF INVENTION: Silk Proteins

; NUMBER OF SEQUENCES: 56

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Birch, Stewart, Kolasch & Birch

; STREET: P.O. Box 747

; CITY: Falls Church

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22040-3487

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/458,298

; FILING DATE: 02-JUN-1995

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/209,747

; FILING DATE: 14-MAR-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Murphy Jr., Gerald M.

; REGISTRATION NUMBER: 28,977

; REFERENCE/DOCKET NUMBER: 1447-104P

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-205-8000

; TELEFAX: 703-205-8050

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2793 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; ORIGINAL SOURCE:

; ORGANISM: Nephila clavipes

; TISSUE TYPE: minor ampullate gland

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 183..2675

; OTHER INFORMATION: /product= "N. clavipes minor

; OTHER INFORMATION: ampullate silk protein"

US-08-458-298-1

Query Match 8.3%; Score 40.4; DB 1; Length 2793;

Best Local Similarity 45.8%; Pred. No. 0.15;

Matches 140; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 167 GAGCCCAAGGAGGAGTGTGGTCCCAAGGAGAGTGTGGTGCATTAATGAGTGGCGGACTGAGGCA 226

Db 520 GTGCGGAGCGGAGCTGTGTCTCTGAGCTGTGGAGCAGGAGCGCGAGGCGGCTGGTG 579

QY 227 GGGGCCAGTGCCACACTGAGCTGTGAGGTGGCCCGAGGCCCGAGAGGAGTGTGGTAC 286

Db 580 GTTACGGTAGAGTGTGGTCTCTGAGCTGTGTGGGCGAGGTGTGAGGGCGCG 639

QY 287 AAGGATGGGAAGAGCTGAGCTCCAGTTCGAAGTGGCATAGAGGCTGCGGGCTGCATG 346

Db 640 GTGGTCTGGATATGGTGGACAAGCGGATATGGTGGCGGAGCAGGAGCTGGTGGCGCTG 699

QY 347 CGGCAGCTGGTGTGGTGCAGCAGGAGCGAGGAGGAGTGTGGGAGGTACACCTGTGAGGCT 406

Db 700 CTGCTGCTGTGGTGCAGGAGCAGGAGGAGTGTGGCGGTTACGGTAGAGGTCTGTGTGGAG 759

QY 407 GGGGGCCAGCGGCTCTCTTCCACCTGGATGTTCAGAGCCCAAGGCGGTGTTTCCAAAG 466

Db 760 CAGGAGCCCTGCGGCTGTGGAGTGGAGGCTACGGTGGTCAAGGTGGGTACGGTGCCG 819

QY 467 GAGCAG 472

Db 820 GAGCAG 825

RESULT 8

US-08-762-106-5

; Sequence 5, Application US/08762106

; Patent No. 5948677

; GENERAL INFORMATION:

; APPLICANT: Jarvik, Jonathan W.

; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE

; TITLE OF INVENTION: TAGGING

; NUMBER OF SEQUENCES: 47

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Harris Brotman

; STREET: 202 Coast Blvd., Suite 111

; CITY: La Jolla

; STATE: California

; COUNTRY: US

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/762,106

; FILING DATE: 09-DEC-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Brotman, Harris F.

; REGISTRATION NUMBER: 35,461

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 654-2428

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1548 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-762-106-5

Query Match 8.2%; Score 39.8; DB 2; Length 1548;

Best Local Similarity 48.2%; Pred. No. 0.19;

Matches 175; Conservative 0; Mismatches 182; Indels 6; Gaps 2;

QY 125 GAGGCTGGGGCCAGAGAGTCTCTTCACTGCATCATCAGAGCCCAAGGCGAGTGT 184

Db 1045 GAGGCGGTGGCGGCGGCCACGTTTGTGTGCGGAGCTCAAGCCCGCGCGGAGC 1104

QY 185 GCCAA---GGAGCAGTGTGGTGCATAATGAGTGGCGACTGAGGAGGGGCCAGTGCCACA 241

Db 1105 GCCGATGCCGTGAGAGCGCGCGCGAGCTGACGGCGCAGCGGAGGAGCGCCCAAC 1164

QY 242 CTGAGCTGTGAGGTGGCCCGCCAGCAGAGAGTGTGTACAGGATGGGAAGAAG 301

Db 1165 GCCAAGTGGAGGGCGGACAAGGCGGAGCGCGCGGAGAGGCGCGCGAGCGGAGCGC 1224



QY 302 CTGAGCTCCAGTTCCAAAGTCGCATAGAGGCTCGGGCTGCATCGCGCAGCTGGTGGT 361  
Db 1225 CGCGGAGGACGAGAGGCTCTCTGGAGAGTTGGCGGCCACGCGCGGAGGCG 1284  
QY 362 CAGCAGGAGCCAGGCGAGATGCTGGGAGTACCTGTGAGGCTGGGGCCAGCGCTC 421  
Db 1285 GAGGAGCGCGGAGGAGCCCGCGGAGCGCGCTGCTGCCGGATGGCGTGGAGC-- 1342  
QY 422 TCCTTCCACCTGGATGTTTCAGAGCCCAAGCGGTGTTTGCAGAGGAGCAGCTGGCAC 481  
Db 1343 -CTGTGACGTGGAGCTGAGTGGCCNAGCGGTGGAGGCTGTGCCNAGCGCGCTC 1401  
QY 482 AGG 484  
Db 1402 AAG 1404

## RESULT 9

US-09-320-774-5  
; Sequence 5, Application US/09320774  
; Patent No. 6265545  
; GENERAL INFORMATION:  
; APPLICANT: Jarvik, Jonathan W.  
; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE  
; TITLE OF INVENTION: TAGGING  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harris Brotman  
; STREET: 202 Coast Blvd., Suite 111  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/320,774  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/762,106  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brotman, Harris F.  
; REGISTRATION NUMBER: 35,461  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 654-2428  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1548 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-09-320-774-5

Query Match 8.2%; Score 39.8; DB 3; Length 1548;  
Best Local Similarity 48.2%; Pred. No. 0.19;  
Matches 175; Conservative 0; Mismatches 182; Indels 6; Gaps 2;  
QY 125 GAGGCTGGGGCCAGAGAGTCTCCTTTCACTGCATCATCAGAGCCCAAGGAGTGT 184  
Db 1045 GAGGCGGTGCGCGCGGCCACGTTTGTGTCGCGAGCTCAAGCCCGGTGGCGAGC 1104  
QY 185 GCCAA---GGAGCAGTTGGTGCAATAGAGTGGCGAGCTGAGGCGAGGCGCCACAGTGCAC 241  
Db 1105 GCGATGCGCTGAGGCGCGCGCGGAGTGACGCGCGAGGAGGCGGCGCCAC 1164

QY 242 CTGAGCTGTAGTGGCCCGAGCCAGAGAGTGCATGTTCAAGGATGGGAAGAG 301  
Db 1165 GCCAAGTGGGAGCGGACAAAGCGAGGCGGCGGAGAGGCGCGCGGAGGCG 1224  
QY 302 CTGAGCTCCAGTTCGAAAGTGGCATAGAGGCTCGGGCTGCATGGCGCAGCTGGTGG 361  
Db 1225 GCGCGGAGGAGCAGAGGCGCTGTGGAGGAGTTGGCGGCCACGCGCGGAGGCG 1284  
QY 362 CAGCAGGAGCCAGGCGAGATGCTGGGAGTACCTGTGAGGCTGGGGCCAGCGCTC 421  
Db 1285 GAGGAGCGCGGAGGAGCCCGCGGAGCGCGCTGCTGCCGGATGGCGTGGAGC-- 1342  
QY 422 TCCTTCCACCTGGATGTTTCAGAGCCCAAGCGGTGTTTGCAGAGGAGCAGCTGGCAC 481  
Db 1343 -CTGTGACGTGGAGCTGAGTGGCCNAGCGGTGGAGGCTGTGCCNAGCGCGCTC 1401  
QY 482 AGG 484  
Db 1402 AAG 1404

## RESULT 10

US-08-762-106-6  
; Sequence 6, Application US/08762106  
; Patent No. 5948677  
; GENERAL INFORMATION:  
; APPLICANT: Jarvik, Jonathan W.  
; TITLE OF INVENTION: READING FRAME INDEPENDENT EPITOPE  
; TITLE OF INVENTION: TAGGING  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harris Brotman  
; STREET: 202 Coast Blvd., Suite 111  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/762,106  
; FILING DATE: 09-DEC-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brotman, Harris F.  
; REGISTRATION NUMBER: 35,461  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 654-2428  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1581 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-762-106-6

Query Match 8.2%; Score 39.8; DB 2; Length 1581;  
Best Local Similarity 48.2%; Pred. No. 0.19;  
Matches 175; Conservative 0; Mismatches 182; Indels 6; Gaps 2;  
QY 125 GAGGCTGGGGCCAGAGAGTCTCCTTTCACTGCATCATCAGAGCCCAAGGAGTGT 184  
Db 1078 GAGGCGGTGCGCGCGGCCACGTTTGTGTCGCGAGCTCAAGCCCGGTGGCGAGC 1137  
QY 185 GCCAA---GGAGCAGTTGGTGCAATAGAGTGGCGAGCTGAGGCGAGGCGCCACAGTGCAC 241



## US-08-938-291A-3

Query Match 8.2%; Score 39.6; DB 3; Length 5020;  
Best Local Similarity 51.1%; Pred. No. 0.3;  
Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 295 GAAGAAGCTGAGCTCCAGTTGAAAGTGGCATAGAGGCTGGGGCTGCATGGCGAGCT 354  
DB 4156 GGAGACCCCGGCTGCACTACGGTGCCTTACTTGGGGCCCGGGCTGCCCGGCTCCA 4097  
QY 355 GGTGGTGCAGCAGGAGCCAGGAGATGTGGGGAGTACACCTGTGAGGCTGGGGGCCA 414  
DB 4096 GGGCGGCAGTCATGGCGGGCCCGAGCGCGGCCGACACTCATGCTGCGTGGCCCC 4037  
QY 415 GCGGCTCTCTCCACTGTGATTTTCAGAGCCCAAGCGGTTGTTCAAGAGAGAGCT 474  
DB 4036 GCTGCTGCGCATCGCGCTGCTGTGTCGCTGCTGTCGCGTGGGGCTGGGGCT 3977  
QY 475 GG 476  
DB 3976 GG 3975

## RESULT 13

US-09-589-619-3/c  
; Sequence 3, Application US/09589619  
; Patent No. 6576442  
; GENERAL INFORMATION:  
; APPLICANT: Lev, Sima  
; Flowman, Gregory D.  
; Schlessinger, Joseph  
; TITLE OF INVENTION: RGB PROTEINS AND RELATED  
; PRODUCTS AND METHODS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRES:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FastSeq  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/589,619  
; FILING DATE: 07-Jun-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/938,291  
; FILING DATE: September 26, 1997  
; APPLICATION NUMBER: 60/027,337  
; FILING DATE: October 11, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 228/172  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5020 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-589-619-3

## RESULT 14

US-09-919-172-30  
; Sequence 30, Application US/09919172  
; Patent No. 6673545  
; GENERAL INFORMATION:  
; APPLICANT: Faris, Mary  
; APPLICANT: Turner, Christopher M.  
; TITLE OF INVENTION: PROSTATE CANCER MARKERS  
; FILE REFERENCE: PA-0036 US  
; CURRENT APPLICATION NUMBER: US/09/919,172  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: 60/222,469  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 102  
; SOFTWARE: PERL Program  
; SEQ ID NO 30  
; LENGTH: 1708  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6673545 1709387CB1  
US-09-919-172-30

Query Match 8.1%; Score 39.4; DB 4; Length 1708;  
Best Local Similarity 47.1%; Pred. No. 0.24;  
Matches 121; Conservative 0; Mismatches 136; Indels 0; Gaps 0;  
QY 6 GTTACAAGGACGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGCATGGAGGTCAGGGCT 65  
DB 629 GGCTCAAGTATGAAATGAGCTGGCCCTCGCCAGGGCTTGGGCTGACATCAACGGCT 688  
QY 66 GCACACGAAGGCTGGTAGTGCAGCAGGTGGGCAAGCAGATGCTGGGGAGTACAGTGG 125  
DB 689 TGGCCGAGTCTCTGGATGAGCTGACCTGGCCAGGACTGACCTGGAGATGCAGATCGAG 748  
QY 126 AGCTGGGGCCAGAGAGTCTCTTTCAACTGCATACAGAGCCCAAGGAGTGTG 185  
DB 749 GCCTGAATGAGGAGCTAGCCTTACCTGAGAAGAACCAAGAGAGAGATGAAGAGTCA 808  
QY 186 CCAAGGAGCAGTTGGTGCATATAGGTGCGGACTGAGGCGAGGGGCGAGTCCACACTGA 245  
DB 809 GCAGCCAGCTGGCGGCCAGGTCAATGTGAGATGACGACGACCCGGGTGTGACCTGA 868  
QY 246 GCTGTGAGTGGCCCCAG 262  
DB 869 CCGTGTGTGGCAGAG 885

## RESULT 15

US-09-010-398-2

```
; Sequence 2, Application US/09010398
; Patent No. 5972615
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: Veltri, Robert W.
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,398
; FILING DATE: Concurrently Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Corder, Timothy S.
; REGISTRATION NUMBER: 38,414
; REFERENCE/DOCKET NUMBER: UROC:015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512-418-3000
; TELEFAX: 512-474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1709 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-010-398-2

Query Match      8.1%; Score 39.4; DB 2; Length 1709;
Best Local Similarity 47.1%; Pred. No. 0.25;
Matches 121; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy      6  GGTACAGGACGGGAAGCGTAGCTCCAGCTCAAAAGTACGCATGGAGGTCAGGGCT 65
Db      636  GGCTCAAGTATGAGAAATGAGCTGGCCCTGCCGCCAGGGCGTTGAGGCTGACATCAACGGCT 695

Qy      66  GCACACGAAGGCTGGTAGTCAGCAGCGTGGCAAGCAGATGCTGGGGAGTACAGCTGG 125
Db      696  TGGCCCGAGTCCTGGATGAGCTGACCTGGCCAGGACTGACCTGGAGATGCAGATCAGG 755

Qy      126  AGGCTGGGGCCAGAGAGTTCCTTTCAACTGCATGCACATCAGAGCCCAAGGGCAGTGTGTTG 185
Db      756  GCCTGAATGAGGAGTACCTGCTGAAGAAAGAACCAAGAGAGAGATGAAGGAGTTCA 815

Qy      186  CCAAGGAGCAGTGGTGCATATAGTGGGAGTGGGAGTGGGCGGCGGCGGCGGCGGCGGCGG 245
Db      816  GCAGCCAGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 875

Qy      246  GCTGTGAGTGGGCCAG 262
Db      876  CCGGTGTGCTGGCAGAG 892
```

Search completed: September 19, 2004, 22:12:35  
Job time : 45.5558 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 19:28:00; Search time 251.035 Seconds  
(without alignments)  
Sequence: 1 gacgtgtacagagcgagga.....ggagcagctggcacacagga 485  
9751.690 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_3662\_4146

Perfect score: 485

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
3: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq.\*  
4: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
5: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq.\*  
6: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq.\*  
7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq.\*  
12: /cgn2\_6/prodata/2/pubpna/US09\_PUB.seq.\*  
13: /cgn2\_6/prodata/2/pubpna/US09\_PUB.seq.\*  
14: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq.\*  
15: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq.\*  
16: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq.\*  
17: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
18: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
19: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	485	100.0	23907	14	US-10-077-130-6
2	485	100.0	24120	14	US-10-077-130-4
3	200.4	41.3	2768	16	US-10-120-988-426
4	185.6	38.3	578	13	US-10-027-632-272054
5	185.6	38.3	578	16	US-10-027-632-272054
6	73.4	15.1	1645	9	US-09-726-643-22
7	73.4	15.1	1645	14	US-10-042-141-22
8	73.4	15.1	2170	16	US-10-264-049-850
9	65.8	13.6	2534	16	US-10-108-260A-1946
10	65.8	13.6	14061	16	US-10-093-463-73
11	65.8	13.6	14061	16	US-10-093-463-71
12	64.2	13.2	7564	17	US-10-476-397-15
13	61.4	12.7	2254	9	US-09-726-643-44
14	61.4	12.7	2254	14	US-10-042-141-44

15	61	12.6	716	10	US-09-822-846-117	Sequence 117, App
16	54.8	11.3	400	13	US-10-085-783A-8373	Sequence 8373, App
17	54.8	11.3	400	16	US-10-242-535A-8373	Sequence 8373, App
18	54.6	11.3	20565	10	US-09-764-891-6045	Sequence 6045, App
19	54.6	11.3	20565	15	US-10-091-438-270	Sequence 270, App
20	54.2	11.2	1266	10	US-09-822-846-116	Sequence 116, App
21	53.6	11.1	2884	16	US-10-094-749-1001	Sequence 1001, App
22	53.2	11.0	9591	10	US-09-764-891-6043	Sequence 6043, App
23	53.2	11.0	9591	15	US-10-091-438-268	Sequence 268, App
24	53.2	11.0	12415	10	US-09-764-891-6044	Sequence 6044, App
25	53.2	11.0	12415	15	US-10-091-438-269	Sequence 269, App
26	49.2	10.1	81940	9	US-09-759-508B-1	Sequence 1, Appli
27	49.2	10.1	81940	10	US-09-980-706-1092	Sequence 1092, App
28	49.2	10.1	81940	10	US-09-873-319-724	Sequence 724, App
29	48.8	10.1	466	15	US-10-029-386-22631	Sequence 22631, A
30	47.2	9.7	3575	17	US-10-641-643-1395	Sequence 1395, App
31	47.2	9.7	3597	14	US-10-098-841-84	Sequence 84, Appl
32	45	9.3	3935	16	US-10-108-260A-534	Sequence 534, App
33	43.4	8.9	671	15	US-10-184-644-346	Sequence 346, App
34	43.4	8.9	671	15	US-10-184-644-346	Sequence 346, App
35	43.4	8.9	11916	15	US-10-156-761-926	Sequence 926, App
36	43.4	8.9	11916	17	US-10-204-862A-3	Sequence 3, Appli
37	43.4	8.9	16596	15	US-10-156-761-930	Sequence 930, App
38	43.4	8.9	30690	17	US-10-204-862A-1	Sequence 1, Appli
39	43.4	8.9	31422	17	US-10-204-862A-2	Sequence 2, Appli
40	43.4	8.9	9025608	15	US-10-156-761-1	Sequence 1, Appli
41	43.4	8.9	9025608	15	US-10-156-761-1	Sequence 1, Appli
42	41	8.5	4936	13	US-10-183-243-14	Sequence 14, Appl
43	41	8.5	5007	15	US-10-307-619-3	Sequence 3, Appli
44	41	8.5	5207	9	US-09-858-664A-1	Sequence 1, Appli
45	41	8.5	5207	13	US-10-697-263-1	Sequence 1, Appli

#### ALIGNMENTS

#### RESULT 1

US-10-077-130-6  
; Sequence 6, Application US/10077130  
; Publication No. US20020168742A1  
; GENERAL INFORMATION:  
; APPLICANT: Kapeller-Libermann, Rosana  
; TITLE OF INVENTION: 59079 and 13599, Protein Kinase Family  
; TITLE OF INVENTION: Members and Uses therefor  
; FILE REFERENCE: MPI2001-047P/IRCP/ (M)  
; CURRENT APPLICATION NUMBER: US/10/077,130  
; PRIOR FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: 60/269201  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 23907  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-077-130-6

Query Match	100.0%	Score	485	DB	14	Length	23907
Best Local Similarity	100.0%	Pred. No.	1.5e-130	Indels	0	Gaps	0
Matches	485	Conservative	0	Mismatches	0	Indels	0
QY	1	GACGTGTACAGGACGGAGAGAGCTGAGCTCCAGTCAAAAGTACGCATGGAGGTCAA	60				
Db	3591	GACGTGTACAGGACGGAGAGAGCTGAGCTCCAGTCAAAAGTACGCATGGAGGTCAA	3650				
QY	61	GGCGTGCACACGAGCGCTGGTAGTCGACAGTGGGCAAGCAGATCTGGGGAGTACAG	120				
Db	3651	GGCGTGCACACGAGCGCTGGTAGTCGACAGTGGGCAAGCAGATCTGGGGAGTACAG	3710				
QY	121	CTGCGAGGTGGGGCCACAGAGTCTCTTTCAATGCACATCAGAGGCCAAGGCAGT	180				
Db	3711	CTGCGAGGTGGGGCCACAGAGTCTCTTTCAATGCACATCAGAGGCCAAGGCAGT	3770				

QY 181 GTTTCACAGGACGAGTTGGTGCATTAATGAGGTGCGGACTGAGGACAGGGGCCAGTGCAC 240  
 |||||  
 Db 3771 GTTTCACAGGACGAGTTGGTGCATTAATGAGGTGCGGACTGAGGACAGGGGCCAGTGCAC 3830  
 |||||  
 QY 241 ACTGAGCTGTGAGGTGGCCAGGCCAGACAGAGGTGACGTGGTACAAAGGATGGAGAA 300  
 |||||  
 Db 3831 ACTGAGCTGTGAGGTGGCCAGGCCAGACAGAGGTGACGTGGTACAAAGGATGGAGAA 3890  
 |||||  
 QY 301 GCTGAGCTCCAGTTCCAAAGTCCGATAGAGCTCCGGGCTGCATGCGGACAGCTGGTGGT 360  
 |||||  
 Db 3891 GCTGAGCTCCAGTTCCAAAGTCCGATAGAGCTCCGGGCTGCATGCGGACAGCTGGTGGT 3950  
 |||||  
 QY 361 GCAGCAGGACAGCCAGGACAGTCTGGGGAGTACACCTGTGAGCTGGGGGCCAGCGGCT 420  
 |||||  
 Db 3951 GCAGCAGGACAGCCAGGACAGTCTGGGGAGTACACCTGTGAGCTGGGGGCCAGCGGCT 4010  
 |||||  
 QY 421 CTCCTTCCACCTGGATGTTTCAGAGCCCAAGGCGGTGTTTGCAAAGGACAGCTGGCACA 480  
 |||||  
 Db 4011 CTCCTTCCACCTGGATGTTTCAGAGCCCAAGGCGGTGTTTGCAAAGGACAGCTGGCACA 4070  
 |||||  
 QY 481 CAGGA 485  
 |||||  
 Db 4071 CAGGA 4075

## RESULT 2

US-10-077-130-4  
 ; Sequence 4, Application US/10077130  
 ; Publication No. US2002016742A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kapeller-Libermann, Rosana  
 ; APPLICANT: Acton, Susan L.  
 ; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family  
 ; FILE REFERENCE: MPI2001-047P1RCP1(M)  
 ; CURRENT APPLICATION NUMBER: US/10/077,130  
 ; PRIOR FILING DATE: 2002-02-15  
 ; PRIOR FILING DATE: 2001-02-15  
 ; NUMBER OF SEQ ID NOS: 9  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 24120  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: 5'UTR  
 ; LOCATION: (1)...(71)  
 ; NAME/KEY: CDS  
 ; LOCATION: (72)...(23978)  
 ; NAME/KEY: 3'UTR  
 ; LOCATION: (23979)...(24120)  
 US-10-077-130-4

Query Match 100.0%; Score 485; DB 14; Length 24120;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-130;  
 Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCTGTGTACAGGACCGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGGAGTCAA 60  
 |||||  
 Db 3662 GAGCTGTGTACAGGACCGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGGAGTCAA 3721  
 |||||  
 QY 61 GGCTGTACACGAGGCTGTGAGTGTGAGCAGGTGGGCAAGCAGATGCTGGGAGTACAG 120  
 |||||  
 Db 3722 GGCTGTACACGAGGCTGTGAGTGTGAGCAGGTGGGCAAGCAGATGCTGGGAGTACAG 3781  
 |||||  
 QY 121 CTCGAGGCTGGGGGCCAGAGAGTCTCTTTTCAACTGCACATCACAGAGCCCAAGGCACT 180  
 |||||  
 Db 3782 CTCGAGGCTGGGGGCCAGAGAGTCTCTTTTCAACTGCACATCACAGAGCCCAAGGCACT 3841  
 |||||  
 QY 181 GTTTCACAGGACGAGTTGGTGCATTAATGAGGTGCGGACTGAGGACAGGGGCCAGTGCAC 240  
 |||||

Db 3842 GTTTCACAGGACGAGTTGGTGCATTAATGAGGTGCGGACTGAGGACAGGGGCCAGTGCAC 3901  
 |||||  
 QY 241 ACTGAGCTGTGAGGTGGCCAGGCCAGACAGAGGTGACGTGGTACAAAGGATGGAGAA 300  
 |||||  
 Db 3902 ACTGAGCTGTGAGGTGGCCAGGCCAGACAGAGGTGACGTGGTACAAAGGATGGAGAA 3961  
 |||||  
 QY 301 GCTGAGCTCCAGTTCCAAAGTCCGATAGAGCTCCGGGCTGCATGCGGACAGCTGGTGGT 360  
 |||||  
 Db 3962 GCTGAGCTCCAGTTCCAAAGTCCGATAGAGCTCCGGGCTGCATGCGGACAGCTGGTGGT 4021  
 |||||  
 QY 361 GCAGCAGGACAGCCAGGACAGTCTGGGGAGTACACCTGTGAGCTGGGGGCCAGCGGCT 420  
 |||||  
 Db 4022 GCAGCAGGACAGCCAGGACAGTCTGGGGAGTACACCTGTGAGCTGGGGGCCAGCGGCT 4081  
 |||||  
 QY 421 CTCCTTCCACCTGGATGTTTTCAGAGCCCAAGGCGGTGTTTGCAAAGGACAGCTGGCACA 480  
 |||||  
 Db 4082 CTCCTTCCACCTGGATGTTTTCAGAGCCCAAGGCGGTGTTTGCAAAGGACAGCTGGCACA 4141  
 |||||  
 QY 481 CAGGA 485  
 |||||  
 Db 4142 CAGGA 4146

## RESULT 3

US-10-120-988-426  
 ; Sequence 426, Application US/10120988  
 ; Publication No. US20030219745A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tang, Y. Tom  
 ; APPLICANT: Goodrich, Ryle  
 ; APPLICANT: Liu, Chenghua  
 ; APPLICANT: Ren, Feiyan  
 ; APPLICANT: Wang, Dunrui  
 ; APPLICANT: Drmanac, Radoje T.  
 ; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and  
 ; FILE REFERENCE: 802CON  
 ; CURRENT APPLICATION NUMBER: US/10/120,988  
 ; CURRENT FILING DATE: 2002-04-11  
 ; PRIOR FILING DATE: 2002-04-11  
 ; PRIOR FILING DATE: 2001-01-30  
 ; NUMBER OF SEQ ID NOS: 441  
 ; SOFTWARE: pt\_FL\_genes Version 2.0  
 ; SEQ ID NO 426  
 ; LENGTH: 2768  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (72)..(2738)  
 US-10-120-988-426

Query Match 41.3%; Score 200.4; DB 16; Length 2768;  
 Best Local Similarity 69.8%; Pred. No. 3.8e-48;  
 Matches 286; Conservative 0; Mismatches 121; Indels 3; Gaps 1;

QY 5 TGGTACAGGACGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGGAGGTCAAGGGC 64  
 |||||  
 Db 2349 TGGTACAGGATGGCATGGAGCTGGGCCACTCCGGTGTGAGCGCTCTTTCAGGAGGATGTG 2408  
 |||||  
 QY 65 TGCACAGC--AAGCTGTGTAGTCACAGAGTGGGCAAGCAGATGCTGGGAGTACAGC 121  
 |||||  
 Db 2409 GGGACGGGCAACCGGCTGGTGGCAGCCACAGTCAACAGGAGGATGAAGGCACCTACTCC 2468  
 |||||  
 QY 122 TGCAGAGCTGGGGGCCAGAGAGTCTCCTTTCAACTGCACATCACAGAGCCCAAGGAGTG 181  
 |||||  
 Db 2469 TGCAGCTGGGCGAGGAGTCTGTGGACTTCGGCTCGCGCTCTTGAGCCCAAGGCGGTG 2528  
 |||||  
 QY 182 TTTTCCAGAGAGCAGTTGGTGCATTAATGAGTGTGGAGTCTGAGCAGGGGCCAGTGCAC 241  
 |||||  
 Db 2529 TTTTCCAGAGAGCAGCGGCTGAGGAGGTGACGCTGAGGTGGGGGCCAGCGCCACG 2588  
 |||||  
 QY 242 CTGAGCTGTGAGGTGGGCCAGGCCCAGACAGAGGTGACGCTGAGTGTGCAAGGATGGGAAG 301  
 |||||

Db 2589 CTGAGCTGTGAGGTGGCCAGGACCATGATGAGGTGACGTGGTACAGGACGGGAAG 2648  
Qy 302 CTGAGCTCCAGTTCGAAGTGGCATAGAGCTGCGGCTGCATGGCGCAGCTGTGGTG 361  
Db 2649 TTGAGCTCCAGCTCGAAGTGCAGTGGAGCTGTGGGCTGTATGCGAGGCTGTGGTG 2708  
Qy 362 CAGCAGGACGAGCCAGGACGATGCTGGGAGGTACACCTGTGAGGCTGGGG 411  
Db 2709 CAGCAGGTGGCCAGGACGATCTCGGATAGTACAGCTGTGAAGCCAGGTG 2758

RESULT 4  
US-10-027-632-272054  
; Sequence 272054, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027.632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 272054  
; LENGTH: 578  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-272054

Query Match 38.3%; Score 185.6; DB 13; Length 578;  
Best Local Similarity 87.5%; Pred. No. 6.2e-44;  
Matches 203; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
Qy 163 CACAGAGCCCAAGGACAGTGTTCCTCAAGGAGCAGTGTGTCATATGAGTGGCGACTGA 222  
Db 347 CCACAGAGCCCAAGTGTGTTTCCCAAGGAGCAGCCACACACAGGAGGTGCAGGCTGA 406  
Qy 223 GGCAGGGCCAGTGCACACTGAGTGTGAGTGGCCCGCCAGCCACAGAGGTGACGTG 282  
Db 407 GCGGGGGCCAGTGCACCGCTGAGTGTGAGTGGCCCGCCAGCCACAGAGGTGACGTG 466  
Qy 283 GTACAGGATGGGAAGCTCAGTCCAGTTCGAAAGTGCATAGAGGTGCGGGCTG 342  
Db 467 GTACAGGATGGGAAGCTCAGTTCGAAAGTGCATAGGTCGCGTGGAGCCGTGGGCTG 526  
Qy 343 CATGCGGCAGCTGGTGTGTCAGCAGGAGCCAGCCAGCAGATGCTGGGGAGTAC 394  
Db 527 CACACGAGGCTGGTGTGTCAGCAGGCGGGCCAGGAGCCGCGGGAGTAC 578

RESULT 5  
US-10-027-632-272054  
; Sequence 272054, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027.632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 272054  
; LENGTH: 578  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-272054  
Query Match 38.3%; Score 185.6; DB 16; Length 578;  
Best Local Similarity 87.5%; Pred. No. 6.2e-44;  
Matches 203; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
Qy 163 CACAGAGCCCAAGGACAGTGTTCCTCAAGGAGCAGTGTGTCATATGAGTGGCGACTGA 222  
Db 347 CCACAGAGCCCAAGTGTGTTTCCCAAGGAGCAGCCACACACAGGAGGTGCAGGCTGA 406  
Qy 223 GGCAGGGCCAGTGCACACTGAGTGTGAGTGGCCCGCCAGCCACAGAGGTGACGTG 282  
Db 407 GCGGGGGCCAGTGCACCGCTGAGTGTGAGTGGCCCGCCAGCCACAGAGGTGACGTG 466  
Qy 283 GTACAGGATGGGAAGCTCAGTCCAGTTCGAAAGTGCATAGAGGTGCGGGCTG 342  
Db 467 GTACAGGATGGGAAGCTCAGTTCGAAAGTGCATAGGTCGCGTGGAGCCGTGGGCTG 526  
Qy 343 CATGCGGCAGCTGGTGTGTCAGCAGGAGCCAGCCAGCAGATGCTGGGGAGTAC 394  
Db 527 CACACGAGGCTGGTGTGTCAGCAGGCGGGCCAGGAGCCGCGGGAGTAC 578

RESULT 6  
US-09-726-643-22  
; Sequence 22, Application US/09726643  
; Patent No. US2002002849A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: 26 Human secreted proteins  
; FILE REFERENCE: P2040P1  
; CURRENT APPLICATION NUMBER: US/09/726,643  
; CURRENT FILING DATE: 2000-12-01  
; PRIOR APPLICATION NUMBER: PCT/US00/15187  
; PRIOR FILING DATE: 2000-06-02  
; PRIOR APPLICATION NUMBER: 60/137,725  
; PRIOR FILING DATE: 1999-06-07  
; NUMBER OF SEQ ID NOS: 190  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 1645  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-726-643-22

Query Match 15.1%; Score 73.4; DB 9; Length 1645;  
Best Local Similarity 49.4%; Pred. No. 2.4e-11;  
Matches 219; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 5 TGGTACAAGACGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGAGGTCAAGGGC 64  
Db 905 TGGTACAAGACGGGACGAGAGTGGAGAGTACTTCGTGGTCTGGAGAAATGAGGGG 964  
QY 65 TGCACACGAAGCTGGTGTAGTCAGCAGGTGGGCAAGAGATGCTGGGAGATACAGTGC 124  
Db 965 CCCCATCGCCGCTGGTGTGTCGCCGCCACCCATCCCTCAGACGGGGCGAGTTTCAGTGC 1024  
QY 125 GAGGCTGGGGCCAGAGAGTCTCCCTTTCAACTGCACATCACAGAGCCCAAGGAGTGT 184  
Db 1025 GTCGCTGGAGATGAGTGTGCCTACTTCACTGTACCAATCACAGAGCTCTCCTGTGATC 1084  
QY 185 GCCAAGGACAGTGGTGTGATTAATGAGTGGGACTGAGGAGGCGGCGAGTGCACACTG 244  
Db 1085 GT---GTATCCAGCGCAAGGTGTATGTGGCAGCGCTCGGCTGGAGCGTGTGTGCTG 1141  
QY 245 AGCTGTGAGGTGGCCAGAGCTCTCCCTTTCAACTGCACATCACAGAGCCCAAGGAGTGT 304  
Db 1142 ACCTGTGAGCTATGCGGCCCTGGCAGAGGTGCGTGGACCAAGGATGGAGAGGTG 1201  
QY 305 AGCTCCAGTTCGAAAGTCCGATAGAGCTCGGGCTGCATCGGCGAGCTGGTGTGTCAG 364  
Db 1202 GTGAGAGCCCGCGCTGCTCTCGAGAAAGACACTGTCCGCCCTGGTGTCTGCC 1261  
QY 365 CAGGACGCCAGGAGATGCTGGGAGTACACTGTGAGGCTGGGGGCCAGCGGTCTCC 424  
Db 1262 GCTGTCCAGCTCGAGGACTCCCGCGAGTCTGTGTGAAATTGACGATGATCGGCTCC 1321  
QY 425 TTCCACCTGGATGTTTCAGAGCC 447  
Db 1322 TTCACTGTCAACGTCACAGAGTC 1344

## RESULT 7

US-10-042-141-22  
; Sequence 22, Application US/10042141  
; Publication No. US20020183503A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: 26 Human secreted proteins  
; FILE REFERENCE: P2040P1  
; CURRENT APPLICATION NUMBER: US/10/042,141  
; CURRENT FILING DATE: 2002-01-11  
; PRIOR APPLICATION NUMBER: 09/726,643  
; PRIOR FILING DATE: 2000-12-01  
; PRIOR APPLICATION NUMBER: PCT/US00/15187  
; PRIOR FILING DATE: 2000-06-02  
; PRIOR APPLICATION NUMBER: 60/137,725  
; PRIOR FILING DATE: 1999-06-07  
; NUMBER OF SEQ ID NOS: 190  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 1645  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-042-141-22

Query Match 15.1%; Score 73.4; DB 14; Length 1645;  
Best Local Similarity 49.4%; Pred. No. 2.4e-11;  
Matches 219; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 5 TGGTACAAGACGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGAGGTCAAGGGC 64  
Db 905 TGGTACAAGACGGGACGAGAGTGGAGAGTACTTCGTGGTCTGGAGAAATGAGGGG 964  
QY 65 TGCACACGAAGCTGGTGTAGTCAGCAGGTGGGCAAGAGATGCTGGGAGATACAGTGC 124  
Db 965 CCCCATCGCCGCTGGTGTGTCGCCGCCACCCATCCCTCAGACGGGGCGAGTTTCAGTGC 1024  
QY 125 GAGGCTGGGGCCAGAGAGTCTCCCTTTCAACTGCACATCACAGAGCCCAAGGAGTGT 184  
Db 1025 GTGCTGGAGATGAGTGTGCTGCTACTTCACTGTCAACATCACAGAGCTCTCCTGTGATC 1084

QY 185 GCCAAGGACAGTGGTGTGATTAATGAGTGGGACTGAGGAGGCGGAGTGCACACTG 244  
Db 1085 GT---GTATCCAGCGCAAGGTGTATGTGGCAGCGCTCGGCTGGAGCGTGTGTGCTG 1141  
QY 245 AGCTGTGAGGTGGCCAGAGCTCCAGAGCCAGAGAGTGCATCGGCGAGTGCATCGGCGAGTGT 304  
Db 1142 ACCTGTGAGCTATGCGGCCCTGGGCGAGAGTGGCTGGACCAAGGATGGAGAGGTG 1201  
QY 305 AGCTCCAGTTCGAAAGTCCGATAGAGCTCGGGCTGCATCGGCGAGTGCATCGGCGAGTGT 364  
Db 1202 GTGAGAGCCCGCGCTGCTCTCGAGAAAGACACTGTCCGCCCTGGTGTGCTGCC 1261  
QY 365 CAGGACGCCAGGAGATGCTGGGAGTACACTGTGAGGCTGGGGGCCAGCGGTCTCC 424  
Db 1262 GCTGTCCAGCTCGAGGACTCCCGCGAGTCTGTGTGAAATTGACGATGATCGGCTCC 1321  
QY 425 TTCCACCTGGATGTTTCAGAGCC 447  
Db 1322 TTCACTGTCAACGTCACAGAGTC 1344

## RESULT 8

US-10-264-049-850  
; Sequence 850, Application US/10264049  
; Publication No. US20040005579A1  
; GENERAL INFORMATION:  
; APPLICANT: Birse et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P4133P1  
; CURRENT APPLICATION NUMBER: US/10/264,049  
; CURRENT FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: PCT/US01/18569  
; PRIOR FILING DATE: 2001-06-07  
; PRIOR APPLICATION NUMBER: US 60/209,467  
; PRIOR FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 4360  
; SOFTWARE: PatentIn Ver. 3.1  
; SEQ ID NO 850  
; LENGTH: 2170  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-264-049-850

Query Match 15.1%; Score 73.4; DB 16; Length 2170;  
Best Local Similarity 49.4%; Pred. No. 2.5e-11;  
Matches 219; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 5 TGGTACAAGACGGGAAGAGCTGAGCTCCAGCTCAAAAGTACGATGAGGTCAAGGGC 64  
Db 1396 TGGTACAAGACGGGCGAGAGGTGGAGAGATGACTTCGTGGTCTGGAGAAATGAGGGG 1455  
QY 65 TGCACACGAAGCTGGTGTAGTCAGCAGGTGGGCAAGAGATGCTGGGAGTACAGTGC 124  
Db 1456 CCCCATCGCCGCTGGTGTGCTGCCGCCACCCAGCCCTCAGACGGGGCGAGTTTCAGTGC 1515  
QY 125 GAGGCTGGGGCCAGAGAGTCTCCCTTTCAACTGCACATCACAGAGCCCAAGGAGTGT 184  
Db 1516 GTCGCTGGAGATGAGTGTGCTACTTCACTGTCAACATCACAGAGCTCTCCTGTGATC 1575  
QY 185 GCCAAGGACAGTGGTGTGATTAATGAGTGGGACTGAGGAGGCGGCGAGTGCACACTG 244  
Db 1576 GT---GTATCCAGCGCAAGGTGTATGTGGCAGCGCTCGGCTGGAGCGTGTGTGCTG 1632  
QY 245 AGCTGTGAGGTGGCCAGGCGCCAGAGAGTGCATCGTGGTACAGAGTGGAGAGAGCTG 304  
Db 1633 ACCTGTGAGCTATGCGGCCCTGGCAGAGTGGCTGGACCAAGGATGGAGAGGTG 1692  
QY 305 AGCTCCAGTTCGAAAGTCCGATAGAGCTCGGGCTGCATCGGCGAGTGTGCTGTCAG 364  
Db 1693 GTGAGAGCGCCCGGCTGCTCCTSCAGAAAGACACTGTCCGCCCGCTGTGTGCTGCC 1752  
QY 365 CAGGACGCCAGGAGATGCTGGGAGTACACTGTGAGGCTGGGGGCCAGCGGTCTCC 424



Db	1753	GCTGTCCAGCTCAGAGACTCGGCGAGTACTTGTGTGAAATTGACGATGAGTCGGCTCC	1812
QY	425	TTCCACCTGGATGTTTCAGAGCC	447
Db	1813	TTCACTGTCAACCGTCACAGAGTC	1835

```

RESULT 9
US-10-108-260A-1946
; Sequence 1946, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108.260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1946
; LENGTH: 2534
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1946

```

	Query Match	13.6%	Score 65.8	DB 16	Length 2534
	Best Local Similarity	51.9%	Pred. No. 4.1e-09		
	Matches 148	Conservative 0	Mismatches 137	Indels 0	Gaps 0
QY	194	CAGTTGTCATATGAGGTGCGGACCTGAGGCAGCGGCCAGTGCACACCTGAGCTGTGAG	253		
Db	256	CTGAAGCGCTGGATGACCTCTCCGACAGAGAGCGGCCACCTGGGCCCTCAGTGTGAA	315		
QY	254	GTGCCCAGGCCACGACAGAGGTGACGTGTACAGGATGGGAGAGAGCTCAGCTCCAGT	313		
Db	316	GTCTCTGACCCCGAGGCCCATGTGGTGTGGCGCAAGATGGCTGCAGCTGGGCCCCAGT	375		
QY	314	TCGAAAGTGCGCATAGAGGCTGCGGGCTGCATGCGGCAGCTGGTGGTGCAGAGCGAGGC	373		
Db	376	GACAAGTATGACTTCTCCACACGCGCGGCCACGCGGGGGCTCGTGGTGCATGAGCTGAGC	435		
QY	374	CAGCAGATGCTGGGGAGTACACCTGTGAGGCTGGGGCCAGCGCTCTCCTTCACACTG	433		
Db	436	CCTGAAGACGCGCGGCTGTACACCTTGCACCTGCGAGTGGGCTCCGAGAGAGACCCGGGCCGGGTC	495		
QY	434	GATGTTTCAGAGCCCAAGCGCGTGTTCACAAAGGAGCAGCTGGCA	478		
Db	496	CGCTGCAGCATCTGCAGCTGGGCATCACCAGAGGCTGGAAGACA	540		

```

RESULT 10
US-10-059-463-73
; Sequence 73, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glennda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerkusen, Bryan
; APPLICANT: Tchervnev, Velizar
; APPLICANT: Gangollis, Esha
; APPLICANT: Vernet, Corinne

```

	Query Match	13.6%	Score 65.8	DB 16	Length 14061
	Best Local Similarity	51.9%	Pred.No. 5.1e-037		
	Matches 148	Conservative 0	Mismatches 137	Indels 0	Gaps 0
QY	194	CAGTTGTGTCNATATAGAGTGC	CGGA	CTGAGCGAGCGGCACGTCC	CACATCTGAGCTGTGAG 253
DB	180	CTGAAAGCGCTGGATGACCTT	TCCG	CAGAGAGCGCGCACCCCTGCC	CTTCAGTGTGNA 239
QY	254	GTGCGCCAGGCCCCACAGAGTG	AGTGC	TGTGTA	CAAGGATGGGHAAGAGTGAGCTCCAGT 313

RESULT 11  
 US-10-093-463-71  
 ; Sequence 71, Application US10093463  
 ; Publication No. US20030208039A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Padigaru, Muralidhara  
 ; APPLICANT: Shenoy, Suresh  
 ; APPLICANT: Kekuda, Ramesh  
 ; APPLICANT: Gusev, Vladimir  
 ; APPLICANT: Pochart, Pascal  
 ; APPLICANT: Zhong, Mei  
 ; APPLICANT: Rastelli, Luca  
 ; APPLICANT: Mezes, Peter  
 ; APPLICANT: Smithson, Glennda  
 ; APPLICANT: Guo, Xiaojia  
 ; APPLICANT: Gerlach, Valerie  
 ; APPLICANT: Casman, Stacie  
 ; APPLICANT: Boldog, Ferenc  
 ; APPLICANT: Li, Li  
 ; APPLICANT: Zerhusen, Bryan  
 ; APPLICANT: Tchernev, Velizar  
 ; APPLICANT: Gangolli, Esha  
 ; APPLICANT: Vernet, Corine  
 ; APPLICANT: Pena, Carol  
 ; APPLICANT: Burgess, Catherine  
 ; APPLICANT: Liu, Xiaohong  
 ; APPLICANT: Spytek, Kimberly  
 ; APPLICANT: Gorman, Linda  
 ; APPLICANT: Spaderna, Steven  
 ; APPLICANT: Voss, Edward  
 ; APPLICANT: Malyankar, Uriel  
 ; APPLICANT: Anderson, David  
 ; APPLICANT: Patturajan, Meera  
 ; APPLICANT: Miller, Charles  
 ; APPLICANT: Taupier, Raymond J. Jr.  
 ; TITLE OF INVENTION: No. US20030208039A1 Antibodies that Bind to Antigenic Polypeptides  
 ; FILE REFERENCE: 21402-290A (Cura 590AT)  
 ; CURRENT APPLICATION NUMBER: US/10/093,463  
 ; CURRENT FILING DATE: 2002-06-24  
 ; PRIOR APPLICATION NUMBER: 60/283,675  
 ; PRIOR FILING DATE: 2001-04-14  
 ; PRIOR APPLICATION NUMBER: 60/338,092  
 ; PRIOR FILING DATE: 2001-12-03  
 ; PRIOR APPLICATION NUMBER: 60/274,281  
 ; PRIOR FILING DATE: 2001-03-08  
 ; PRIOR APPLICATION NUMBER: 60/274,101  
 ; PRIOR FILING DATE: 2001-03-08  
 ; PRIOR APPLICATION NUMBER: 60/325,681  
 ; PRIOR FILING DATE: 2001-09-27  
 ; PRIOR APPLICATION NUMBER: 60/304,354  
 ; PRIOR FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: 60/279,995  
 ; PRIOR FILING DATE: 2001-03-30  
 ; PRIOR APPLICATION NUMBER: 60/294,899  
 ; PRIOR FILING DATE: 2001-05-31  
 ; PRIOR APPLICATION NUMBER: 60/287,424



```
; ORGANISM: Homo sapiens
US-10-042-141-44

Query Match      12.7%; Score 61.4; DB 14; Length 2254;
Best Local Similarity 49.2%; Pred. No. 7.6e-08;
Matches 218; Conservative 0; Mismatches 221; Indels 4; Gaps 2;

QY 5 TGGTAAAGACGCGGGAAGCTGAGCTCCAGCTCAAAAGTACGATGGAGGTCAAGGGC 64
Db 1400 TGGTAAAGACGCGGAGAGGTGGAGAGTGTCTCGTGGTCTCGAGATCAGGGG 1459
QY 65 TGCACACGAGCTGGTAGTCAGACAGAGTGGGCAAGCAGATGCTGGGAGATCAGCTGC 124
Db 1460 CCCCATCGCGCTCGTGTGCGCCGCCACCCAGCCCTCAGACGGGGCGAGTTTCAGTGC 1519
QY 125 GAGGCTGGGGGCGAGAGATCTCTCTTTCAACTGCACATCAGAGGCCCAAGGCAGTGT 184
Db 1520 GTGCTGGAGATGAGTGTGCTTACTTCACTGTCAACATCAGAGCTCTCTCTGTGATC 1579
QY 185 GCCAAGGACAGTGTGTCGATATAGAGTGGGACTGAGGCGGGGCCAGTGCACACTG 244
Db 1580 GT---GTATCCAGCGCAAGGTGTATGTGGCAGCCGTGCGCCCTGGAGCGCTGTGCTG 1636
QY 245 AGCTGTGAGGTGGCCCGGCGCCAGACAGAGGTGACGTGTACAAAGATGGGAAGACTG 304
Db 1637 ACCTGTGAGTATGCGGCCCTTGGCAGAGTGTGCTGGACCAAGATGGAGAGAGTG 1696
QY 305 AGCTCCAGTTTGAAGTCCGATAGAGCTCGGGCTGCATGCGGCGAGCTGGTGGTGCAG 364
Db 1697 GTGGAGACCCCG-CGCTGCTCTCTGCGAAGAAAGACACTGTCCGCCCGCTGTGCTGCC 1755
QY 365 CAGGCGAGCCAGGCGAGATGCTGGGGAGTACACCTGTGAGGCTGGGCGCCAGCGCTTCC 424
Db 1756 GCTGTGAGCTCGAGGACTCCGGCGAGTACTTGTGTGAATTTGACGATGAGTGGCGCTCC 1815
QY 425 TTCCACCTGGATGTTTCAGAGCC 447
Db 1816 TTCACTGTCACTCCACAGAGTC 1838

Search completed: September 20, 2004, 02:51:55
Job time : 261.035 secs
```

```
RESULT 15
US-09-822-846-117
; Sequence 117, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Treacy, Maurice
; APPLICANT: Merberg, David
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6400
; CURRENT APPLICATION NUMBER: US/09/822,846
; PRIORITY FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,605
; PRIORITY FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 117
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: September 19, 2004, 15:07:04 ; Search time 1414.09 Seconds  
(without alignments)  
10242.013 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_3662\_4146  
Perfect score: 485  
Sequence: 1 gactgtgtacaaaggacggga.....ggagcagctggcacacagga 485

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hcc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hcc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gss1:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	342.2	70.6	632	14	CF180572
C 2	314.8	64.9	621	14	CF179218
C 3	216	44.5	1142	12	B680164
C 4	212.8	43.9	401	10	AW898791

5	207.6	42.8	522	28	AQ595824
6	207.2	42.7	591	13	BU439885
7	194.8	40.2	658	28	AZ383233
C 8	193.8	40.0	574	9	AI595154
9	180.4	37.2	786	14	CD099595
10	161.2	33.2	588	13	BU436339
C 11	142.4	29.4	343	10	BE159678
C 12	140.6	29.0	373	9	AA061364
13	132.4	27.3	631	12	BG793386
14	132.2	27.3	397	28	B2837763
15	127.8	26.4	1192	28	CC247513
16	112.6	23.2	179	12	BG989628
17	112.4	23.2	472	29	CB495069
18	85.8	17.7	724	29	CNS03WVR
C 19	84.4	17.4	875	29	CNS03WVR
20	75.6	15.6	470	10	BF087837
21	74	15.3	612	13	BU947303
C 22	73.4	15.1	525	10	BF087734
23	73.4	15.1	886	13	BU535739
C 24	73.4	15.1	982	13	BQ877554
25	73.4	15.1	1085	12	BM546177
26	73.4	15.1	1098	12	BM543727
27	73.4	15.1	1201	13	BX377136
28	73.4	15.1	2066	29	AY399870
29	72.4	14.9	505	10	BF087860
C 30	72	14.8	805	13	BX381923
31	71.8	14.8	575	13	BU072543
32	71.6	14.8	569	12	BM254228
C 33	71.4	14.7	982	13	BX343487
C 34	71	14.6	533	10	AW995362
35	70.4	14.5	4130	11	AK048144
36	70.4	14.5	5558	11	AK029863
37	67.8	14.0	555	9	AU119815
38	67.8	14.0	724	9	AU141131
C 39	67.8	14.0	1201	13	BX377135
40	67.8	14.0	1201	13	BX377135
41	67	13.8	649	14	CF262757
42	66	13.6	743	10	BE314256
43	66	13.6	835	10	BE260056
44	66	13.6	949	13	BQ645083
45	66	13.6	949	13	BQ839538

ALIGNMENTS

RESULT 1  
CF180572/c  
LOCUS 815696 MARC 3PIG Sus scrofa cDNA 3', mRNA sequence.  
DEFINITION CF180572.1 GI:33292348  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CF180572 632 bp mRNA linear EST 28-JUL-2003  
815696 MARC 3PIG Sus scrofa cDNA 3', mRNA sequence.  
CF180572.1 GI:33292348  
EST.  
Sus scrofa (pig)  
Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
1 (bases 1 to 632)  
Smith, T.P.L., Freking, B.A., Ford, J.J., Vallet, J.L., Fox, J., Wise, T.A., Noneman, D.J., Wray, J.E. and Keele, J.W.  
A second set of porcine ESTs from a pooled-tissue normalized library  
Unpublished (2003)  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@mail.marc.usda.gov  
Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim\_alt option. Vector identified with cross\_match v0.990329.  
Plate: SRG8009 row: O column: 4

US-10-077-130-4\_COPY\_3662\_4146

```

Seq primer: TAGAAGGCACAGTCGAGG.
Location/Qualifiers
1..632
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 3P1G"
/note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including brain, liver, muscle, placenta/endometrium,
ovary, testes, and bone marrow."

ORIGIN
Query Match 70.6%; Score 342.2; DB 14; Length 632;
Best Local Similarity 81.8%; Pred. No. 4e-65;
Matches 395; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 1 GACGTGTACAGGACGCGGAGAGAGCTGAGCTCCAGCTCAAAAGTACGATGGAGGTCAA 60
Db 512 GACTTGGTACAGGATGGGAAGAGCTGAGTTCAGGCTCAAAAGTGGGTGGAGGCCAA 453

QY 61 GGGCTGCACACGAAGGCTGTAGTCAGCAGGTGGGCAAGCAGATGCTGGGAGTACAG 120
Db 452 GGGCTGCACACGAAGGCTGTAGTCAGCAGGTGGGCAAGCAGATGCTGGGAGTACAG 393

QY 121 CTGGAGGCTGGGGCCAGAGAGTCTCTTTCACTGCAATCAAGAGCCCAAGGAGT 180
Db 392 CTGGAGGCTGGGGCCAGAGAGTCTCTTTCACTGCAATCAAGAGCCCAAGGAGT 333

QY 181 GTTTCGCAAGGACAGTGTGGTGCATAATGAGTGGGACTGAGGAGGCGGCGAGTGCAC 240
Db 332 GTTTCGCAAGGACAGTGTGGTGCATAATGAGTGGGACTGAGGAGGCGGCGAGTGCAC 273

QY 241 ACTGAGCTGTAGGTGGCCAGGCCAGCAGAGGTGACGTGTGATCAAGGATGGGAAGAA 300
Db 272 AATGAGCTGTAGGTGGCCAGGCCAGCAGAGGTGACGTGTGATCAAGGATGGGAAGAA 213

QY 301 GCTGAGCTCCAGTTCGAAAGTGGCCATAGAGGCTGCGGGTGCATGCGGAGCTGTGTGT 360
Db 212 GCTGAGCTCCAGTTCGAAAGTGGCCATAGAGGCTGCGGGTGCATGCGGAGCTGTGTGT 153

QY 361 GCAGAGCGAGGCGAGCAGATGCTGGGGAGTACACTGTGAGGCTGGGGCCAGCGGCT 420
Db 152 GCAGAGCGGCGAGCAGATGCTGGGGAGTACACTGTGAGGCTGGGGCCAGCGGCT 93

QY 421 CTCCTTCCACTGGATCTTTCAGAGCCCAAGCGGCTGTTTGCACAGGAGCAGTGCACA 480
Db 92 CTCCTTCCACTGGATCTTTCAGAGCCCAAGCGGCTGTTTGCACAGGAGCAGTGCACA 33

QY 481 CAG 483
Db 32 CAG 30

RESULT 2
CF179218
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Sus scrofa (pig)
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 621)
AUTHORS
Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J.,
Wise,T.A., Nonneman,D.J., Wray,J.E. and Keele,J.W.
TITLE
A second set of porcine ESTs from a pooled-tissue normalized
library
JOURNAL
Unpublished (2003)

Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross match v0.990329.
Plate: SRG8009 row: O column: 4
Seq primer: GTAATACGACTCAGTATAGG.
Location/Qualifiers
1..621
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 3P1G"
/note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including brain, liver, muscle, placenta/endometrium,
ovary, testes, and bone marrow."

ORIGIN
Query Match 64.9%; Score 314.8; DB 14; Length 621;
Best Local Similarity 80.1%; Pred. No. 4.4e-59;
Matches 370; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 1 GACGTGTACAGGACGCGGAGAGAGCTGAGCTCCAGCTCAAAAGTACGATGGAGGTCAA 60
Db 160 GACTTGGTACAGGATGGGAAGAGCTGAGTTCAGGCTCAAAAGTGGGTGGAGGCCAA 219

QY 61 GGGCTGCACACGAAGGCTGTAGTCAGCAGGTGGGCAAGCAGATGCTGGGAGTACAG 120
Db 220 GGGCTGCACACGAAGGCTGTAGTCAGCAGGTGGGCAAGCAGATGCTGGGAGTACAG 279

QY 121 CTGGAGGCTGGGGCCAGAGAGTCTCTTTCACTGCAATCAAGAGCCCAAGGAGTGCAC 180
Db 280 CTGGAGGCTGGGGCCAGAGAGTCTCTTTCACTGCAATCAAGAGCCCAAGGAGTGCAC 339

QY 181 GTTTCGCAAGGACAGTGTGGTGCATAATGAGTGGGACTGAGGAGGCGGCGAGTGCAC 240
Db 340 GTTTCGCAAGGACAGTGTGGTGCATAATGAGTGGGACTGAGGAGGCGGCGAGTGCAC 399

QY 241 ACTGAGCTGTAGGTGGCCAGGCCAGCAGAGGTGACGTGTGATCAAGGATGGGAAGAA 300
Db 400 ACTGAGCTGTAGGTGGCCAGGCCAGCAGAGGTGACGTGTGATCAAGGATGGGAAGAA 459

QY 301 GCTGAGCTCCAGTTCGAAAGTGGCCATAGAGGCTGCGGGTGCATGCGGAGCTGTGTGT 360
Db 460 GCTGAGCTTCAGCTCAAAAGTGGCGTGTGGAGGCCAATGGCTGCACCCGGATGCTGTGT 519

QY 361 GCAGAGCGAGGCGAGCAGATGCTGGGGAGTACACTGTGAGGCTGGGGCCAGCGGCT 420
Db 520 GCATCATGCGGGCAATGCATATGCTGGGGAGTACAGTGCAGTGCAGTGCAGTGCATATGT 579

QY 421 CTCCTTCCACTGGATCTTTCAGAGCCCAAGCGGCTGTTTGC 462
Db 580 CTCCTTCCACTGGATCTTTCAGAGCCCAAGCGGCTGTTTGC 621

RESULT 3
BG680164/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
BG680164
602626828F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4751671 5',
mRNA sequence.
BG680164
BG680164.1 GI:13911561
EST.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 1142)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: James Cleaver, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: InCyte Genomics, Inc.

DNA distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLAM10608 row: h column: 08

High quality sequence stop: 708.

Location/Qualifiers

1. 1142

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4751671"

/tissue\_type="squamous cell carcinoma"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_lib="NCI CGAP SKN4"

/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.5kb. Library constructed by Life Technologies. Note: this is a NCI CGAP Library."

FEATURES

source

```

ORIGIN
Query Match 44.5%; Score 216; DB 12; Length 1142;
Best Local Similarity 80.8%; Pred. No. 3.5e-37;
Matches 252; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 155 CTGCACATCAGAGAGCCCAAGGCGAGTGTTCACAGGACAGTTGGTGCATATCAGGTG 214
    |||
Db 322 CTGCTCTCCCCAGAGCCCAAGGCGGTGTTCCTAAGGAGCAGCCACACAGGGAGGTG 263
    |||
QY 215 CGGACTGAGGCAGGGGCCAGTGCCACACTGAGCTGTGAGTGTGCCCCAGGCCACAGACAG 274
    |||
Db 262 CAGGCTGAAGCAGGGGCTAGGCCACACTGAGCTGTGAGTGTGCCCCACAGCACAG 203
    |||
QY 275 GTGACGTGGTTCAAAGGATGGGAAGAGCTGAGTCCAGTTCGAAAGTCGCAGTACAGGCT 334
    |||
Db 202 GTGACTTGGTTCAAAGGACGGGAAGAGCTGAGCTCCAGCTTGAAGTGCATGTGGAGCT 143
    |||
QY 335 CGGGGCTGCATGCGGCAGCTGGTGGTGCAGCAGGAGGCGCAGGCAGATGCTGGGAGGTAC 394
    |||
Db 142 GCAGGCTGTACACGGAGGCTGGTGGTGCAGCAGGAGGCGCAGCAGACACCGGAGAGTAT 83
    |||
QY 395 ACCTGTGAGGCTTGGGGGCCACGGGCTCTCCCTTCACACTGGATGTTTTCAGAGCCCCAAGCG 454
    |||
Db 82 AGCTGTGAGGCTGGGGGCCACAGCAGCTCTCCTTCGGCTTCAGTGGCAGGTCACTGGTGT 23
    |||
QY 455 GTGTTTGCAAG 466
    |||
Db 22 GGGGATGCTGAG 11
    |||

```

RESULT 4	AW898791	401 bp	linear	EST 24-MAY-2000
LOCUS	CM0-NN0075-130400-332-f06	NN0075	Homo sapiens cDNA, mRNA sequence:	
DEFINITION	AW898791			
ACCESSION	AW898791.1	GI:8062996		
VERSION				
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 401)			
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S. Briones,M.R.,			

Nagai, M.A., da Silva W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsumura, A., Bata, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20203663  
PUBMED 10737800  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL:  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl1=st2=CMO-NN0075-130400-332-f06&tl3=2000-04-13&tl4=1>)

FEATURES	source
High quality sequence	401.
Location/Qualifiers	
1. .401	
/organism="Homo sapiens"	
/mol_type="rRNA"	
/db_xref="taxon:9606"	
/dev_stage="Adult"	
/clone_lib="NN0075"	
/notes="Organ: nervous normal; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 136,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."	
ORIGIN	
Query Match	43.9%; Score 212.8; DB 10; Length 401;
Best Local Similarity	79.1%; Pred. No. 11e-36;
Matches 253; Conservative	0; Mismatches 67; Indels 0; Gaps 0;
155	CTGCACATCAGAGCCCAAGGAGGTGTTGTCACAGGACGATGTGTGTCATTAATAGAGTG 214
Db	61 CTGCTTTCCCCAGAGGCCAAGCGGTGTTGTCACAGGACGACGACGAGCGAGGAGGTG 120
QY	215 CGGACTGAGCGAGGGCCAGTGCCACATCGAGCTGTGAGGTGGCCAGGCCACACACAG 274
Db	121 CAGCTGAGCGGGGACCATGCCCAGCTGAGCTGCGAGGTGGCCAGGCCACACACAG 180
QY	275 GTGACGTGTTACAAGATCGGAAGAGCTGAGCTCCAGTTCGAAAGTCGCATAGAGGCT 334
Db	181 GTGACGTGTTACAAGGACGGGAAGAACTGAGCTCCAGTTCGAAAGTCGGAATGAGGCC 240
QY	335 GCGGGCTGCATGCGGCGAGGTGTGTGTGCACAGCGAGGCCACGAGCATGCTCGGGGAGTAC 394
Db	241 GTGGGCTGCACACGAGGCTGTGTGTGTGCAGGAGGCAGGCCAGCGCGGGGAGTAC 300
QY	395 ACCTGTGAGGCTGGGGGCCAGCGGCTCTCTTCCACCTGGATGTTTCAGAGCCCAAGCG 454
Db	301 AGCTGCAAGCCCGGGATCAGCGGTGTCCTTCCAGCTGACGTGCGAGGTCAGTGTGTTT 360
QY	455 GTGTTTCAAAGGAGCAGCT 474
b	361 GGAGGTCTCTGAGGAGCTCT 380

## ORIGIN

Query Match	43.9%	Score	212.8	DB	10	Length	401
Best Local Similarity	79.1%	Pred. No.	1.1e-36				
Matches	253	Conservative	0	Mismatches	67	Indels	0
						Gaps	0
QY	155	CTGCACATCACAGAGCCCAAGCAGCGTGTTCGACAGGAGCAGTTGGTGCATCAATGAGGTG	214				
Db	61	CTGCCTTCCCAAGCCCAAGCGGGTGTTCGACAGGACGCCAGCGAGCGAGGAGGTG	120				
QY	215	CGGACTGAGGCAGGGGCCAAGTGCACACTGAGCTGTGAGGTGGCCCCAGGCCCCAGACAGAG	274				
Db	121	CAGCTGAGCGGGGACCAAGTGCACCGCTGAGCTCGAGGTGGCCCCAGGCCCCAGACAGAG	180				
QY	275	GTGACGTGGTACAAAGGATGGGAAGACTGAGCTCCAGTTCGAAAGTCGGCATGAGGCT	334				
Db	181	GTGACGTGGTACAAAGGACGGGAAGAACTGAGCTCCAGTTCGAAAGTCGGAATGGAGCC	240				
QY	335	GCGGCTGCATCGGCAGCTGTGGTGCAGCAGGACGAGCCAGGCAGATGCTGGGGAGTAC	394				
Db	241	GTGGGCTGTCACACGGAGGCTGTGGTGCAGGAGGACAGCCACGGCGGAGGTAC	300				
QY	395	ACCTGTGAGGCTGGGGGCCAGCGGCTCTCCTTCCACCTGGATGTTTCAGAGCCCAAGCGC	454				
Db	301	AGCTGCAAGCCCGGGGATCAGCGGCTGTCTCTCCAGCTGCACGTGGCAGGTCAGTGTCTTT	360				
QY	455	GTGTTTCCAAAGAGCAGCT	474				
Db	361	GGAGTCTCTGAGGAGCTCT	380				

[illegible]





Db 353 CCTCAGCTGTGAAGTAGCAGGATACACCGAAGTAAATGGTGTACAGGATGAAGACT 412  
 QY 301 GCTCAGCTCCAGTTTCGAAGTGCATAGAGGCTGCGGCTGTGATCGGCGAGCTGGTGGT 360  
 Db 413 GCTTGCCTCTCTAGAAAGTTCAAAATGGAACTGTGGCAAAACCCGGCTCTGGTTGT 472  
 QY 361 GCAGCGCAGGCGCAGCAGATGCTGGGAGTACACCTGTGAAGCTGGGGCGCAGCGGCT 420  
 Db 473 GGAGCAGCTGGAGAAGAAGATGCTGGAGAATACGCTGTGAGGCTCCCGGCAGAGGCT 532  
 QY 421 CTCCTTCCACTGATGTTTCAGAGCCCAAGGCGGTGTTGCAAAAGGA 468  
 Db 533 GACCTTCAAGTTGGAAGCAACTGAACAGAGGCTAAATTTGAAAAGGA 580

RESULT 7  
 AZ383233  
 LOCUS  
 DEFINITION IM0140117R Mouse 658 bp DNA linear GSS 02-OCT-2000  
 clone UUGC1M0140117 R, genomic survey sequence.

ACCESSION  
 AZ383233  
 VERSION  
 1  
 KEYWORDS  
 GSS.  
 SOURCE  
 Mus musculus (house mouse)

ORGANISM  
 Mus musculus

REFERENCE  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (Bases 1 to 658)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,  
 Reilly, M., Rose, R., Rose, R., Stokes, R., Tingey, A., von  
 Niederhausern, A., and Wright, D., Weiss, R.,

TITLE  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

JOURNAL  
 COMMENT  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0140 row: L column: 17  
 Seq primer: CACACAGGAACACCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 658.

FEATURES  
 source  
 1. .658  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0140117"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PWD42 (gi|4732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

# ORIGIN

Query Match 40.2%; Score 194.8; DB 28; Length 658;  
 Best Local Similarity 81.3%; Pred. No. 1.3e-32;  
 Matches 226; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 166 AGAGCCCAAGGCGAGTGTTCACAGGAGCAGTGTGTGATATGAGTGGGACTGAGGC 225  
 Db 182 AGACACCAAGCTGTATGTTTCCACAGGAGCAGCAGCAGGAGTGAAGCTGAGGC 241  
 QY 226 AGGGCCCAAGTGCACACTGAGCTGTGAGTGGCCAGCCAGAGGAGTGAAGTGA 285  
 Db 242 GGGGAACAGTGCACCTGAGCTGCGAGTGGCCAGCCAGCAGTGAAGTGAAGTGT 301  
 QY 286 CAAGATGGAGAGAGCTGAGCTCCAGTTCGAACTGCGCATAGAGGCTGGGGTGCAT 345  
 Db 302 CAAGACCGGAAGAAGCTGAGCTCGAGTTCGAACTGCGGAGGCTCGGGTGTCTC 361  
 QY 346 GCGGCGCAGCTGTGTGTCAGCAGGCGCAGGAGGAGTGTGGGAGTACACCTGTGAGGC 405  
 Db 362 CAGGAGCTGTGTGTGTCAGCAGGCGGCAAGCGGAGTGTGGGAGTACAGCTGCGAGGC 421  
 QY 406 TGGGGCCAGCGCGCTCTCTCCACTGATGTTTCAG 443  
 Db 422 CGGGGCTCAGAAAGCTCTCTCCGCTGGACGTGGCAG 459

# RESULT 8

AI595154/c  
 LOCUS

DEFINITION

AI595154

ACCESSION

AI595154

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

REFERENCE

AUTHORS

1 (bases 1 to 574)

Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
 Person, B., Swaller, I., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
 Ritter, E., Kohn, S., Shin, T., Jackson, J., Cardenas, M., McCann, R.,  
 Waterston, R. and Wilson, R.

The WashU-NCI Mouse EST Project 1999

Unpublished (1999)

TITLE

JOURNAL

COMMENT

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

This read is a RESEQUENCE of a previously sequenced mouse clone

This read has been verified (found to hit its original self in the

correct orientation)

MGI:308283

Seq primer: -40RP from Gibco

High quality sequence stop: 418

POLYA=No.

Location/Qualifiers

1. .574

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="CD-1"

/db\_xref="taxon:10090"

/clone="IMAGE:514435"

```

/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse testis (#937308)"
/note="Organ: testis; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

```

## ORIGIN

```

Query Match 40.0%; Score 193.8; DB 9; Length 574;
Best Local Similarity 80.9%; Pred. No. 2e-32;
Matches 225; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 166 AGAGCCCAAGGACAGTGTTCGCAAGGAGCAGTTCGTCATATGAGGTGGGACTGAGGC 225
DB |||
DB 465 AGACACAGGCTGATGTTGCAAGGAGCAGCAGGACGACCGAGGTGAGGCTGAGGC 406
QY 226 AGGGCCAGTCCACACTGAGCTGTGAGGTGGCCAGGCCACAGAGAGTGTGAGTGA 285
DB |||
DB 405 GGGGAACAGTCCACCTGAGCTCGAGGTGGCCAGGCCAGACTGAGGTGACATGGTT 346
QY 286 CAAGGATGGGAAGAGTGTGAGTCCAGTTCGAAAGTGGCAGAGGCTGGGCTGCAT 345
DB |||
DB 345 CAAGACGGGAAGAGTGTGAGTCCAGTCCAGGTGCGGTGGAGGCTTCGCTC 286
QY 346 CGGCAGCTGTGTGAGCAGGAGGACGAGGAGTGTGGGAGTACACCTGTGAGGC 405
DB |||
DB 285 CAGGAGGCTGTGTGAGCAGGAGGAGGAGGAGTGTGGGAGTACAGTTCGAGGC 226
QY 406 TGGGGCCAGGCGCTCCCTCCACCTGGATGTTTCAG 443
DB |||
DB 225 CGGGGGCAGAAAGTCTCCTTCGCGCTGGAGCTGGCAG 188

```

```

RESULT 9
LOCUS CD099595 786 bp mRNA linear EST 15-MAY-2003
DEFINITION AGNCOURT_13997453 NICHD_XGC_Tadi Xenopus laevis cDNA clone
IMAGE:6940598 5', mRNA sequence.
ACCESSION CD099595
VERSION CD099595.1 GI:30752698
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 786)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Drs. Donald Brown and Liqun Cai
CDNA Library Preparation: CLONTECH
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM3267 row: e column: 13
High quality sequence stop: 437.
Location/Qualifiers
1..786
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clones="IMAGE:6940598"
/dev_stage="metamorphosis stage 53"

```

```

/clone_lib="NICHD XGC Tadi"
/note="Organ: Developing Tadpole; Vector: pDNR-LIB;
Site 1: Sfi; Site 2: Sfi; 5' and 3' adaptors were used in
cloning as follows: 5' adaptor sequence:
5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGGCGGCGGCGGACATG-dt(30)EN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.6 kb
(range 0.9-3.0 kb). 15/15 colonies contained inserts by
PCR. This library was enriched for full-length clones and
was constructed by Clontech Laboratories (Palo Alto, CA)."

```

## ORIGIN

```

Query Match 37.2%; Score 180.4; DB 14; Length 786;
Best Local Similarity 61.1%; Pred. No. 2.1e-29;
Matches 292; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

QY 5 TGCTACAAAGGCGGGAAGCTGAGCTCCAGCTCAAAGTACGATGGAGGTCAAGGC 64
DB |||
DB 38 TGCTACAAAGGAGGAACTGATCACCTCCAGCAAGAAATGAGGTAGAGTCTAAGGC 97
QY 65 TGCACACGAAGCTGTGTAGTGCAGCAGGTGGCCAAAGCAGATGTGGGAGTACAGTGC 124
DB |||
DB 98 AAATCTCGTTCCTGGTGGAGAAAGTAGAGAAACAAGATGCCGAGAGTACAGCTGT 157
QY 125 GAGGCTGGGGCCAGAGAGTCTCTTCAACTGCACATCACAGAGCCCAAGGAGTGT 184
DB |||
DB 158 GAGGCTGGGGCCAGAGAGTCAATTTAAAGTAATGTTAAAGAACCTGAGCCGGCATTC 217
QY 185 GCCAAGGAGCAGTGTGTGTGCATATGAGTGGCGACTGAGGCGAGGGGCCAGTGC 244
DB |||
DB 218 ACCAATACAGACAAGATACAGAGAGGTTGAGGCCACCTTAATGAAATGCCACATG 277
QY 245 AGCTGTGAGGTGGCCAGGCCCCAGACAGAGTGCAGTGTGTACAGGATGGGAAGCTG 304
DB |||
DB 278 AGTTGTGAGGTGTCCCGAGGCAAGCTGAGGTGAGGTGGTACAAAGGAAGGAACTGATA 337
QY 305 AGCTCCAGTTCGAAAGTGGCATAGAGGCTCGGGGCTGCATCGCGGAGCTGGTGTGCAG 364
DB |||
DB 338 ACCTCCACAGAGAGTGCAGGTGAGTCTGAAGGGAATCTCCCGCTGTGTGTGAG 397
QY 365 CAGGAGCCCGGAGCAGATGTGGGGAGTACACCTGTGAGGCTGGGGCCAGCGCTCTCC 424
DB |||
DB 398 AAAGTGGAGAAACAAGATGTTGGAGATATACAGCTGTGAGGCTGCCGGGCAAAAGATCAAT 457
QY 425 TTCACCTGGATGTTTTCAGAGCCCAAGGCGGTGTTTTCAGAGGAGGAGCTGGGCACACA 482
DB |||
DB 458 TTAAAGTTACAGTAAAGAAATCTGAACCAAGCTTCCACCAAGGATAGGTACAGA 515

```

```

RESULT 10
LOCUS BU436399 588 bp mRNA linear EST 29-NOV-2002
DEFINITION 604145537F1 CSEQRBN11 Gallus gallus cDNA clone CHES198519 5', mRNA
sequence.
ACCESSION BU436399
VERSION BU436399.1 GI:25925710
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 588)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
TITLE JOURNAL
MEDLINE 2233534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)

```

PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.  
Location/Qualifiers

FEATURES

source  
1..588  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="Layer and broiler"  
/db\_xref="taxon:9031"  
/clone="CHEST98519"  
/sex="Male and female"  
/tissue\_type="muscle"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="CSEB9N11"  
/note="Vector: pBluescript II KS(+); Site 1: EcoRI;  
Site 2: NotI; This normalized library was constructed from  
1 million independent clones. cDNA synthesis was initiated  
using an oligo(dT) primer, using methylated C in the first  
strand synthesis reaction. Following this first strand  
reaction, double-stranded cDNA was blunted, ligated to  
NotI adapters, digested with EcoRI, size-selected, and  
cloned into the NotI and EcoRI compatible sites of a  
custom modified MCS of the pBluescript (KS+) vector. The  
library was normalized in 2 rounds using conditions  
adapted from Soares et al., PNAS (1994) 91: 9228-9232 and  
Bonaldo et al., Genome Research 6 (1996): 791, except that  
a significantly longer reannealing hybridization was  
used."

ORIGIN

Query Match 33.2%; Score 161.2; DB 13; Length 588;  
Best Local Similarity 67.1%; Pred. No. 3.1e-25;  
Matches 243; Conservative 0; Mismatches 118; Indels 1; Gaps 1;  
Qy 77 CTGTAGTGCAGCAGGTGGGCAAGCAGATCTCTGGGAGTACACTCGAGGCTGGGGC 136  
Db 1 CTGGTGGTGGAGCCAGGTGGAGAAAGATGC-AGGGAGTACACCTCGAGGCTGGTGGC 59  
Qy 137 CAGAGAGTCTCTTCAACTGCACATCACAGAGCCCAAGGAGGTGTTCGCAAGGACAG 196  
Db 60 CAGAACTGACCTTCAAGTACTGTCTCAGAACGAGGATGTTTATCATCACAGAG 119  
Qy 197 TTGTTGCAATATGAGTGGCGAAGTACAGCAGGCGCCAGTGCCACACTGAGCTGTAGGTG 256  
Db 120 AAGGTGCAGAGGAGGTGAAGGCTGCACCAACAGAAATGCCAGCTGAGTGGGAGGTG 179  
Qy 257 GCCAGGCCACAGAGGTGACGTGTACAGAGTGGAGAGAGCTGAGCTCCAGTTTCG 316  
Db 180 GGCAGGAGAGACCGAGGTGAAGTGTGTACAGAGGAGGAGAACTGATCATCTCAGCAAG 239  
Qy 317 AAAGTGGCATAGAGGCTGCGGGTGTGATGCGGAGCTGTGTGTGTCAGCAGGAGGCCAG 376  
Db 240 AAGTTCAGGTTGAGTACAGAGGCAAACTGCTGGCTGTGTGTGAGCCAGTGGAGAAG 299  
Qy 377 GCAGATGCTGGGAGTACACCTGTGAGCTGGGGCCAGCGGCTCTCTTCCACTGGAT 436  
Db 300 AGAATGCGAGGGAGTACACCTGCGAGGCTGCTGGCCAGAAAGTGAAGCTTCAAGATTACT 359  
Qy 437 GT 438  
Db 360 GT 361

RESULT 11

BE159678/c  
LOCUS BE159678 343 bp mRNA linear EST 21-JUN-2000  
DEFINITION MRO-HT0407-180400-015-e07 HT0407 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE159678  
VERSION BE159678.1 GI:8622399  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (Bases 1 to 343)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,  
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
10737800  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the PAPESP/LICR Human Cancer Genome  
Project, this entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2=MR0-HT0407-180  
400-015-e07&t3=2000-04-18&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 343.  
Location/Qualifiers

TITLE

JOURNAL  
MEDLINE  
PUBMED  
COMMENT

FEATURES

1..343  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="HT0407"  
/note="Organ: head neck; Vector: puc18; Site 1: SmaI;  
Site 2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

ORIGIN

Query Match 29.4%; Score 142.4; DB 10; Length 343;  
Best Local Similarity 86.2%; Pred. No. 3.4e-21;  
Matches 169; Conservative 0; Mismatches 26; Indels 1; Gaps 1;  
Qy 163 CACAGAGCCCAAGGAGTGTTCGCAAGGAGCAGTGTGTGCATAATGAGGTGCGACTGA 222  
Db 228 CCAGAGCCCAAGGTGTGTTCGCAAGGAGCAGCAGCAGGAGGTGCAGGCTGA 169  
Qy 223 GGCAGGGCCAGTGCACACTGAGCTGTGAGTGGCCAGGCCAGCCAGACAGAGTGCAGCTG 282  
Db 168 GCGCGGGCCAGTGCACGCTGAGCTGCAGGTGGCCAGGCCAGGCCAGACAGAGTGCAGTG 109  
Qy 283 GTACAGAGTGGGAAGAAGCTGAGCTCCAGTTCGAAAGTGGCAGATAGAGGCTGCGGGCTG 342  
Db 108 GTACAGAGTGGGAAGAAGCTGAGTTCAGAGTCCGAAGTGGCGGTGA-CCCGTGGGCTG 50  
Qy 343 CATGCGGCAGCTGGTG 358  
Db 49 CACACGAGGCTGGTG 34

RESULT 12

AA061264/c  
LOCUS AA061264 373 bp mRNA linear EST 03-FEB-1997  
DEFINITION ml39h10.r1 Stratagene mouse testis (#937308) Mus musculus cDNA  
clone IMAGE:514435 5', similar to TR:G1017425 G1017425  
IPISKPLPKVTLSDRGVPLKATMEFNTAENITINLKSVTADAGRYETTAANSSTGTTKAFINI  
VVLDPRGPPT

GPVVISDITESVTLKWEPKYDGGSOVNYILLKRETSTAVTVEVSATVARTMKVMKL . . . ;  
 mRNA sequence.  
 AA061284  
 VERSION AA061284.1 GI:1555080  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus (house mouse)

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 373)  
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseestowatson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:308283  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -28ml3 rev1: ET from Amersham  
 High quality sequence stop: 366.  
 Location/Qualifiers  
 1..373  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="CD-1"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:514435"  
 /sex="males"  
 /tissue\_type="testis"  
 /dev\_stage="10-12 week old"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /clone\_lib="Stratagene mouse testis (#937308)"  
 /note="Organ: testis; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dr. Average insert size: 1.0 Kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN  
 Query Match 29.0%; Score 140.6; DB 9; Length 373;  
 Best Local Similarity 84.5%; Pred. No. 8.8e-21;  
 Matches 15; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
 257 GCCAGCCGACAGAGGTCAGTGGTACAGGATGGGAAGAGTCCAGTCCAGTTCG 316  
 Db GCCAGCCGACAGTGAAGTGACATGGTTCAAGAGCGGAGAGTGAAGTCCAGTTCG 314  
 QY 317 AAAGTGGCATAGAGGTCGGGCTGATGCGGACCTGGTGGTGAGCAGGCGCAG 376  
 Db 313 AAGTGGCGTGAAGGCTCGGCTCTCCAGGAGCTGGTGGTGAGCAGGCGGCGCAG 254  
 QY 377 GCAGATGCTGGGAGTACACTGTGAGGCTGGGGGCGCAGCGGCTCTCCACCTGGAT 436  
 Db 253 GCGGATGCTGGGAGTACAGTCCGAGGCGGCGGCGCAGAGCTCTCTCCGCTGGAC 194  
 QY 437 GTTTCAG 443  
 Db 193 GTGGCAG 187

RESULT 13  
 BG793386  
 LOCUS 631 bp mRNA linear EST 16-MAY-2001

DEFINITION UTWSM SM11F11 UTWS Adult Mouse Skeletal Muscle Library Mus musculus cDNA clone UTWS SM11F11, mRNA sequence.  
 ACCESSION BG793386  
 VERSION BG793386.1 GI:14128956  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus (house mouse)

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 631)  
 Gallardo, T.D., Schageman, J.J., Pertsemlidis, A., Garner, H.R., Williams, R.S., and Shchet, R.V.  
 TITLE UT Southwestern Medical Center, Adult Mouse Skeletal Muscle cDNA Library  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Schageman JJ  
 Shohet/Garner Labs  
 University of Texas Southwestern Medical Center  
 6000 Harry Hines Blvd., NA2.226, Dallas, TX 75390, USA  
 Tel: 214 848 1674  
 Email: Jeff.Schageman@UTSouthwestern.edu  
 cDNA library constructed by UTWS as a component of the Program for Genomic Applications (PGA) and the Reynolds Heart Disease Prevention Grants for use in cDNA microarray experiments. Sequence Quality: Sequence ends were trimmed based on percentage of ambiguous base calls or 'N's in windowed segments. Sequencing: First-pass sequencing; ABI Prism 377 sequencer and analysis software.  
 Seq primer: M13/pUC Reverse.  
 Location/Qualifiers  
 1..631  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="UTWSM SM11F11"  
 /sex="Pooled"  
 /tissue\_type="Diaphragm/Hind limb muscles"  
 /cell\_type="Skeletal muscle"  
 /dev\_stage="2 months"  
 /lab\_host="DH5a"  
 /clone\_lib="UTWS Adult Mouse Skeletal Muscle Library"  
 /note="Vector: pAMP10 (Gibco); Cloned unidirectionally. Primer: Oligo dr. RNA Isolation: cytoplasmic RNA preps (Manniat); Cloning Technique: CUA Cloning (CloneAmp, Life Technologies); Average insert size: 1.8 Kb; Insertion site: TAGTCACATGATCTGAGT---3'. Other information regarding entire library may be found at http://pga.swmed.edu/Data/libraries/microarray\_cdna\_libraries.htm."

ORIGIN  
 Query Match 27.3%; Score 132.4; DB 12; Length 631;  
 Best Local Similarity 60.6%; Pred. No. 7.2e-19;  
 Matches 231; Conservative 0; Mismatches 149; Indels 1; Gaps 1;  
 QY 1 GACCTGGTACAGACCGGAGAGCTGAGCTCCAGCTCAAAAGTACCGATGAGGTCAA 60  
 Db 73 GACTTGTCTGAAGATGGTCTGAGCTGCTCCCTAGGGCCCAAGTATGAGATGAGGTGTC 132  
 QY 61 GGGCTGCACACGAGAGGCTGGTGTGTCAGCAGGTGGGCAAGCAGATGCTGGGAGGTACAG 120  
 Db 133 AGCTGGGAGCGGAGCACTGCTGGTGGGAGCTGGGACAGATGACGCTGGCCTCTATGA 192  
 QY 121 CTGGAGGCTGGGGCCAGAGAGTCTCTTTTCACTGCACTCACAGAGCCCAAGGCAGT 180  
 Db 193 GTGTGTGAGTCTGGGAGCGGCACTGCTTACCAAGCTGTTGGTCAAGAGCCCAAGATGGT 252  
 QY 181 GTTTGCCAAGGAGCAGTGTGGTGCATATGATGAGTGGCAGCTGAGGCGGAGGCGGAGTCCAC 240  
 Db 253 GTTTGCCAAGGCGCAGCAGCAGCAGCAGCAGTGGTGAAGGCGAGGCGGCGGCGGCGGCGG 312  
 QY 241 ACTGAGCTGTGAGGTGGCCCGAGCCCGCAGCAGAGGTGAGTGGTGTACAAGATGGGAAG-A 299  
 Db 313 ACTGAGCTGTGAGGTGGCCCGAGCCCGCAGCAGCAGCAGTGGTGTACAAGATGGGAAGAA 372

QY 300 AGCTGAGCTCCAGTTGCAAGTGGCATAGAGGCTGGGGCTGCATCGCGCAGCTGTGG 359  
 Db 373 AGCTGAGCTCCAGTTCGAGGTGCGCGTGGAGGCTCGGGGTTCTTCCRANGAAGC 432  
 QY 360 TGCAGCAGGCGAGCGCAGCAG 380  
 Db 433 TGNTTGGGTGCAACAAGGCCG 453

RESULT 14  
 BZ837763 397 bp DNA linear GSS 18-MAR-2003  
 LOCUS CH240\_248L16.TV CHORI-240 Bos taurus genomic clone CH240\_248L16,  
 DEFINITION genomic survey sequence.  
 ACCESSION BZ837763  
 VERSION BZ837763.1 GI:29065122  
 KEYWORDS GSS.  
 SOURCE Bos taurus (cow)  
 ORGANISM Bos taurus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.  
 1 (bases 1 to 397)  
 AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,  
 Shvartsbeyn, A., Gebregorgis, B., Chen, D., Riggs, F., de Jong, P.,  
 Crawford, A.M., and McEwan, J.C.  
 TITLE Bovine BAC End Sequences from Library CHORI-240  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Shaving Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org

Clones are derived from the bovine BAC library CHORI-240  
 (http://www.chori.org/bacpac/bovine240.htm). For BAC library  
 availability, please contact Pieter de Jong (pdejong@mail.choi.org).  
 Clones may be purchased from BACPAC Resources  
 (http://www.chori.org/bacpac/ordering/information.htm). This work  
 was undertaken as part of the International Bovine BAC Mapping  
 Consortium (IBBMC) by AgResearch Ltd., New Zealand and The  
 Institute of Genomic Research (TIGR), USA.  
 Plate: 248 row: L column: 16  
 Seq primer: 17  
 Class: BAC ends.  
 Location/Qualifiers  
 1..397  
 /organism="Bos taurus"  
 /mol\_type="genomic DNA"  
 /strain="breed: Hereford"  
 /db\_xref="taxon:9913"  
 /clone="CH240\_248L16"  
 /sex="Male"  
 /cell\_type="Blood"  
 /clone\_lib="CHORI-240"  
 /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;  
 Hereford bull Li Domino 99375; CHORI-240 Bovine BAC  
 library (Bull) produced by Pieter de Jong"

Query Match 27.3%; Score 132.2; DB 28; Length 397;  
 Best Local Similarity 80.3%; Pred. No. 6.5e-19;  
 Matches 155; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 163 CACAGAGCCCAAGGAGCTGTTGCCAAGAGCAGTGTGTCATATAGAGTCCGACTGA 222  
 Db 205 CCCAGAGCCCTCGGTGGTGTTCGCCAAGAGCAGCGCCAGCAGTGAAGTGCAGGCCGT 264  
 QY 223 GGCAGGGGCGAGTGCACACTGAGCTGTGAGGTGCCAGGCCAGGAGGAGTGCAGCTG 282  
 Db 265 GGCAGGACCACTGTCACACTGAGTGTGAGGTGCCAGGCCAGGAGGAGTGCAGTGTG 324

QY 283 GTACAAGGATGGGAAGAGCTGAGCTCCAGTTCGAAAGTGGCCATAGAGGCTCGGGCTG 342  
 Db 325 GTACAAGGAGCGCAAGAGCTGAGTTCGAGCTCGAAAGTGGCTGTGGAGGCCACGGCGC 384  
 QY 343 CATCGGCGAGCTG 355  
 Db 385 CGGGCGCGGCTG 397

RESULT 15  
 CC247513 1192 bp DNA linear GSS 12-MAY-2003  
 LOCUS CH261-127G1\_RM1.1 CH261 Gallus gallus genomic clone CH261-127G1,  
 DEFINITION genomic survey sequence.  
 ACCESSION CC247513  
 VERSION CC247513.1 GI:30574176  
 KEYWORDS GSS.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 1 (bases 1 to 1192)  
 AUTHORS Krenitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,  
 Warren, W., Graves, T., Mardis, E. and Wilson, R.  
 TITLE Gallus gallus BAC End Reads  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Richard K. Wilson  
 Genome Sequencing Center  
 Washington University School of Medicine  
 Email: submissions@watson.wustl.edu  
 Insert Length: 182000 Std Error: 0.00  
 Seq primer: RM1 TACGACTCACTATAGGAGAG  
 Class: BAC ends  
 High quality sequence start: 40  
 High quality sequence stop: 596.  
 High quality sequence length: 1192

Location/Qualifiers  
 1..1192  
 /organism="Gallus gallus"  
 /mol\_type="genomic DNA"  
 /strain="Red Jungle Fowl"  
 /db\_xref="taxon:9031"  
 /clone="CH261-127G1"  
 /sex="female"  
 /cell\_line="UCD001, inbred 256"  
 /clone\_lib="CH261"  
 /note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;  
 CH261 Female Chicken library - for library and clone  
 ordering information: http://www.chori.org/bacpac"

Query Match 26.4%; Score 127.8; DB 28; Length 1192;  
 Best Local Similarity 65.7%; Pred. No. 1e-17;  
 Matches 186; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 156 TGCACATCAGAGCCCAAGGAGCTGTTGCCAAGAGCAGTGTGTCATATAGAGTGC 215  
 Db 450 TTCCAATTCAGGAAGCACAGAGGATGCTTTATCAACAAGGAGAAGGTGCAGAGGAGTGA 509  
 QY 216 GGAAGTGGAGGCGGCGCAGTGCACACTGAGTGTGAGTGTGCCAGGCCAGGAGGAGG 275  
 Db 510 AGGCTGCACCAACAGAAAATGCCACGCTGAGTGTGAGGAGGAGGAGGAGGAGGAGG 569  
 QY 276 TGACGTGTGTTACAAGGATGGGAAGAGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 335  
 Db 570 TGAAGTGTGTTACAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 629  
 QY 336 CGGCTGTGATCGCGCAGCTGTGTTGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 395  
 Db 630 AGGGCAAACTGCGTCCGCTGGTGTGAGCCAGTGTGAGCAGGAGGAGGAGGAGGAGGAG 689  
 QY 396 CCTGTGAGGCTGGGGCGCAGCGGCTCTCTTCCACCTGGATGT 438

Wed Sep 22 12:33:12 2004

us-10-077-130-4\_copy\_3662\_4146.rst

Page 10

Db 690 CCTGCGAAGCTGCTGGCCCAAAACTGACCTTCCAAATTACTGT 732  
||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: September 19, 2004, 22:07:25  
Job time : 1418.09 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 11:04:42 ; Search time 444 Seconds  
(without alignments)  
5070.582 Million cell updates/sec

Title: US-10-077-130-5  
Perfect score: 41273  
Sequence: 1 MDQPFSGAPRFLTRPKAFV.....RNREKRALLYKSHNLQAVR 7968

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : A\_Geneseq\_29Jan04.\*
- 1: Geneseq1980s.\*
  - 2: Geneseq1990s.\*
  - 3: Geneseq2000s.\*
  - 4: Geneseq2001s.\*
  - 5: Geneseq2002s.\*
  - 6: Geneseq2003as.\*
  - 7: Geneseq2003bs.\*
  - 8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41273	100.0	7968	6	ABG76187 Human ser
2	13710	33.2	2630	6	ABG76186 Human ser
3	13528	32.8	2596	4	AB30569 A splice
4	10519.5	25.5	4691	5	ABP70084 Human NOV
5	10506	25.5	4675	5	ABP70085 Human NOV
6	9834.5	23.8	2328	6	ABP58227 Human cel
7	8423	20.4	1665	5	AA015372 Human myo
8	8410	20.4	1665	5	AA015372 Human myo
9	8407	20.4	1618	4	AA015372 Human myo
10	8403	20.4	1610	4	AA015372 Human myo
11	7073	17.1	1351	4	AB30567 Amino aci
12	4533	11.0	871	4	AA015370 A splice
13	4529	11.0	871	5	AA015370 A splice
14	3878.5	9.4	31267	6	ABG74786 Human kin
15	3580.5	8.7	26926	4	AAU05396 Human tit
16	2447.5	5.9	548	4	AB30571 A full le
17	2415.5	5.9	548	4	AB30572 A full le
18	2360	5.7	4412	3	AA053666 Sequence
19	2328.5	5.6	3267	7	ADC99075 Human KPP
20	2324	5.6	3208	7	ABG74766 Human NOV
21	2319	5.6	3268	7	ABG79959 Human kin
22	2270	5.5	3252	7	ABG74767 Human NOV
23	2203	5.3	3262	7	ABG79961 Mouse ser
24	2135	5.2	2380	5	AA019160 Human kin
25	2127	5.2	2286	4	AA065635 Novel pro

26	2102	5.1	5635	5	ABP60991	Abp60991 Novel hum
27	2085.5	5.1	3186	7	AD47672	Ad47672 Human NOV
28	2067	5.0	2231	7	AD879962	Ad879962 Human KIA
29	1840	4.5	3331	6	ABU07377	Abu07377 Human pro
30	1763.5	4.3	4796	4	AB58665	Ab58665 Drosophil
31	1752	4.2	4495	6	ABU69135	Abu69135 Human NOV
32	1683	4.1	335	5	ABP05662	Abp05662 Human ORF
33	1553	3.8	1237	7	ADD14123	Add14123 Human src
34	1487.5	3.6	7107	4	AB58144	Ab58144 Drosophil
35	1249	3.0	3117	3	AAV53667	Aav53667 Sequence
36	1112	2.7	2173	5	ABP69251	Abp69251 Human pol
37	1081	2.6	4393	4	AB311889	Ab311889 Amino aci
38	1081	2.6	4436	4	ABG23265	Abg23265 Novel hum
39	1074	2.6	4391	6	AA34390	Aa34390 Human per
40	1072	2.6	209	3	AA040558	Aa040558 Human ORF
41	1072	2.6	209	5	ABP31663	Abp31663 Human ORF
42	1025	2.5	1953	5	AAU84351	Aau84351 Protein M
43	1018	2.5	5701	4	AB36684	Ab36684 Peptide #
44	1018	2.5	5701	4	AB22021	Ab22021 Protein #
45	1018	2.5	5701	4	AA069843	Aa069843 Human bon

ALIGNMENTS

RESULT 1  
ABG76187 standard; protein; 7968 AA.  
XX  
AC ABG76187;  
XX  
XX  
DT 09-MAY-2003 (first entry)  
XX  
DE Human serine/threonine or protein kinase 12599.

XX Human; enzyme; serine/threonine kinase; protein kinase; 12599;  
KW cardiovascular disease; heart failure; myocardial infarction;  
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma; immunogen;  
KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;  
KW haemolytic anaemia; cellular proliferative disorder; cancer;  
KW protein kinase disorder; autoimmune disorder; diabetes mellitus;  
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;  
KW multiple sclerosis.  
XX  
XX Homo sapiens.  
XX  
XX US2002168742-A1.  
XX  
XX 14-NOV-2002.  
XX  
XX 15-FEB-2002; 2002US-00077130.  
XX  
XX 15-FEB-2001; 2001US-0269201P.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Kapeller-Libermann R, Acton SL;  
XX  
XX WPI; 2003-298729/29.  
XX  
XX N-PSDB; ABX11642.  
XX  
XX Novel isolated human protein kinase, designated 59079 or 12599  
XX polypeptide, useful as diagnostic and therapeutic agents for preventing  
XX cardiovascular diseases, proliferative disorders, and protein kinase  
XX disorders.  
XX  
XX Claim 8; Page 84-104; 119pp; English.  
XX  
XX The invention relates to an isolated human serine/threonine or protein  
XX kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule  
XX comprising at least 85% identity to the nucleic acids appearing as  
XX ABX11641 and ABX11642 or their complement, a naturally occurring variant  
XX of the kinases or their fragments. Also included are a non-human host

125

cell containing the nucleic acids, an antibody specific for the proteins, identifying a compound which binds to the kinase (by contacting the kinase or a cell expressing the kinase with a test compound and determining whether the kinase binds to the test compound) and modulating the activity of kinase using the identified compound. The kinases and their encoding nucleic acids are useful as diagnostic and therapeutic agents for preventing a disease or condition associated with an aberrant or unwanted 59079 or 12599 activity in a subject, including cardiovascular diseases such as heart failure, and myocardial infarction; disorders involving blood vessels such as atherosclerosis, and Kaposi's sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia, Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders such as cancer; and protein kinase disorders such as autoimmune disorders, diabetes mellitus, psoriasis, inflammatory bowel disease, rheumatoid arthritis, and multiple sclerosis (many examples of diseases and disorders are included in the specification). The kinases, their encoding nucleic acids and antibodies are useful in screening assays, detection assays (e.g. forensic biology), and predictive medicine (e.g. diagnostic assays, prognostic assays, and monitoring clinical trials and pharmacogenomics). The kinases and their encoding nucleic acids are useful as query sequences to perform a search against public databases to identify other family members or related sequences. The present sequence represents kinase 12599

Sequence 7968 AA;

Query Match 100.0%; Score 41273; DB 6; Length 7968;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 7968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDQPFSGAPFLTRPFAFVSVGKDATLSQIVGNPTPOVSWKDDQOQPTAGARFLAQ 60  
 DB 1 MDQPFSGAPFLTRPFAFVSVGKDATLSQIVGNPTPOVSWKDDQOQPTAGARFLAQ 60

QY 61 DGDLYRLTILDALGDSQYVCRARNAIGFAFAVGLQVDAEACAEQAPHLRPTSTR 120  
 DB 61 DGDLYRLTILDALGDSQYVCRARNAIGFAFAVGLQVDAEACAEQAPHLRPTSTR 120

QY 121 VREGSEATFCRVGSRPAPVSWKGRRLGEPDGPVRVVEELGEASALRIARPRDGG 180  
 DB 121 VREGSEATFCRVGSRPAPVSWKGRRLGEPDGPVRVVEELGEASALRIARPRDGG 180

QY 181 TYEVAENPLGAASAAALVVDSDAATASRPGTSTAALLAHLQRRREARAEAPASPP 240  
 DB 181 TYEVAENPLGAASAAALVVDSDAATASRPGTSTAALLAHLQRRREARAEAPASPP 240

QY 241 STGTRCTCTEGKHARLSCVYTGPKPETVWKDGLVTEGRRHVYVEDAQENFVLKILF 300  
 DB 241 STGTRCTCTEGKHARLSCVYTGPKPETVWKDGLVTEGRRHVYVEDAQENFVLKILF 300

QY 301 CKQSDRGLYTCTASNLGQTVSSVLVVRPAPVFPFKRLQDLEVRKESATFLCEVPQPS 360  
 DB 301 CKQSDRGLYTCTASNLGQTVSSVLVVRPAPVFPFKRLQDLEVRKESATFLCEVPQPS 360

QY 361 TEAAWFKEETRLMASAKYHIEEBGTERRLTVRNVSADDDAVYICETPEGSRTVAEALVQ 420  
 DB 361 TEAAWFKEETRLMASAKYHIEEBGTERRLTVRNVSADDDAVYICETPEGSRTVAEALVQ 420

QY 421 NLLRKLPRKTRAVRGDTAMFCVELAVPVPVHVRNQEVEVAGRVASAEGRTHLTITS 480  
 DB 421 NLLRKLPRKTRAVRGDTAMFCVELAVPVPVHVRNQEVEVAGRVASAEGRTHLTITS 480

QY 481 QCCLVEDGVAFMAGDCQSTFRFCVSAAPRPPLOPPVDPVVKARMSSSVILSNPPHGE 540  
 DB 481 QCCLVEDGVAFMAGDCQSTFRFCVSAAPRPPLOPPVDPVVKARMSSSVILSNPPHGE 540

QY 541 RPVTIDGVLVEKKLGYTWIRCHEAEWATPELTVDADVAEENGFQFVSALNSFGQSPY 600  
 DB 541 RPVTIDGVLVEKKLGYTWIRCHEAEWATPELTVDADVAEENGFQFVSALNSFGQSPY 600

QY 601 LEFPGTVHLAPKLAVRTPKAVQAVEGGEVTFSDLTVASAGBWFLDGQALKASSVYEIH 660  
 DB 601 LEFPGTVHLAPKLAVRTPKAVQAVEGGEVTFSDLTVASAGBWFLDGQALKASSVYEIH 660

QY 661 CDTRHTLTIREVPASLHGAQLKFVANGIESSIRMEVEAAAPGLTANKPFAAAAREVLARL 720  
 DB 661 CDTRHTLTIREVPASLHGAQLKFVANGIESSIRMEVEAAAPGLTANKPFAAAAREVLARL 720

QY 721 HEEAQLLAELSQQAAAVTWLKDGRITLSPGPKYEVQASAGRRLVLRDVARDADAGLYECVS 780  
 DB 721 HEEAQLLAELSQQAAAVTWLKDGRITLSPGPKYEVQASAGRRLVLRDVARDADAGLYECVS 780

QY 781 RGRITAYQLSVQGLARFLHKDMAGSCVDVAGGPAQFECETSEAHVHVHWYKDMELGHS 840  
 DB 781 RGRITAYQLSVQGLARFLHKDMAGSCVDVAGGPAQFECETSEAHVHVHWYKDMELGHS 840

QY 841 GERFLQEDVGTGRHRLVAATVTRQDEGTYSYCRVGSDDVDFRLRVSEPKVVFVFAKEQLARRKL 900  
 DB 841 GERFLQEDVGTGRHRLVAATVTRQDEGTYSYCRVGSDDVDFRLRVSEPKVVFVFAKEQLARRKL 900

QY 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVMCEATCTRRLLVVOQAGADAGEY 960  
 DB 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVMCEATCTRRLLVVOQAGADAGEY 960

QY 961 SCEAGGQRLSHFLDYKEPKVVFVFAKQDQVAHSEVQAEAGANATLSCEVAQAQAEVVMYKDGK 1020  
 DB 961 SCEAGGQRLSHFLDYKEPKVVFVFAKQDQVAHSEVQAEAGANATLSCEVAQAQAEVVMYKDGK 1020

QY 1021 KLSSSLKVHVEAKGCRRLVVOQAGKTADGYSCEARGQVSRFLRHITTEPKMFAKQSV 1080  
 DB 1021 KLSSSLKVHVEAKGCRRLVVOQAGKTADGYSCEARGQVSRFLRHITTEPKMFAKQSV 1080

QY 1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSKVMCEATCTRRLLVVOQAGAD 1140  
 DB 1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSKVMCEATCTRRLLVVOQAGAD 1140

QY 1141 AGEYSCEAGGQVSRFLHITTEPKGVFAKEQSVHNEVQAEAGTAMLSCEVAQAQTEVTWY 1200  
 DB 1141 AGEYSCEAGGQVSRFLHITTEPKGVFAKEQSVHNEVQAEAGTAMLSCEVAQAQTEVTWY 1200

QY 1201 KDGKLSKSSKVMCEATCTRRLLVVOQAGADAGEYSCEAGGQVSRFLHITTEPKVFAK 1260  
 DB 1201 KDGKLSKSSKVMCEATCTRRLLVVOQAGADAGEYSCEAGGQVSRFLHITTEPKVFAK 1260

QY 1261 EQLVHNEVTRTAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVMCEATCTRRLLVVOQ 1320  
 DB 1261 EQLVHNEVTRTAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVMCEATCTRRLLVVOQ 1320

QY 1321 GQADAGEYTCAGGQRLSHFLDYSEPKVFAKEQLAHRKVKQAEAGATATLSCEVAQAQTE 1380  
 DB 1321 GQADAGEYTCAGGQRLSHFLDYSEPKVFAKEQLAHRKVKQAEAGATATLSCEVAQAQTE 1380

QY 1381 VTWYKDGKLSKSSKVMCEATCTRRLLVVOQAGADAGEYSCEAGGQRLSHFLDYSEPKV 1440  
 DB 1381 VTWYKDGKLSKSSKVMCEATCTRRLLVVOQAGADAGEYSCEAGGQRLSHFLDYSEPKV 1440

QY 1441 VFAKEQVHREVEVQAGAGASTTILSCEVAQAQTEVVMYKDGKLSKSSKVMCEATCTRLV 1500  
 DB 1441 VFAKEQVHREVEVQAGAGASTTILSCEVAQAQTEVVMYKDGKLSKSSKVMCEATCTRLV 1500

QY 1501 VQAGQADAGEYSCEAGSQRSLSHFLHVAEPKAVFAKEQPASREVEQAEAGTATLSCEVAQ 1560  
 DB 1501 VQAGQADAGEYSCEAGSQRSLSHFLHVAEPKAVFAKEQPASREVEQAEAGTATLSCEVAQ 1560

QY 1561 AQTEVTWYKDGKLSKSSKVMCEATCTRRLLVVOQAGADAGEYSCEAGGQRLSHFLHVA 1620  
 DB 1561 AQTEVTWYKDGKLSKSSKVMCEATCTRRLLVVOQAGADAGEYSCEAGGQRLSHFLHVA 1620

QY 1621 EPKVVFAKEQPAHREVEQAEAGASTLSCEVAQAQTEVTWYKDGKLSKSSKVMCEATCT 1680  
 DB 1621 EPKVVFAKEQPAHREVEQAEAGASTLSCEVAQAQTEVTWYKDGKLSKSSKVMCEATCT 1680

QY 1681 RRLVVOQAGADAGEYSCEAGGQRLSHFLHVAEELPOISERPCCRREPLVVKHEHDIILTA 1740  
 DB 1681 RRLVVOQAGADAGEYSCEAGGQRLSHFLHVAEELPOISERPCCRREPLVVKHEHDIILTA 1740



QY 1741 TLATPSAATVTLKDGVEIRSKHETASQDTHTLVHGAQVLDLSAISCRVGAEGQDF 1800  
Db 1741 TLATPSAATVTLKDGVEIRSKHETASQDTHTLVHGAQVLDLSAISCRVGAEGQDF 1800  
QY 1801 PVQVEEAAKFCRLLEPVCGELGTVTLACELSPACAEVVMRCNTOPRVGKRFQWVAEG 1860  
Db 1801 PVQVEEAAKFCRLLEPVCGELGTVTLACELSPACAEVVMRCNTOPRVGKRFQWVAEG 1860  
QY 1861 PVRSITVLGLRAEDAGVVCESRDHDSAOITVSPVVKFMSGLSTVVAEEGGEATFQC 1920  
Db 1861 PVRSITVLGLRAEDAGVVCESRDHDSAOITVSPVVKFMSGLSTVVAEEGGEATFQC 1920  
QY 1921 VVSPSDVAVVMFRDGCALLQPSKFAISQSGASHSLTISDLVLEDAGQITVEAEGASSAA 1980  
Db 1921 VVSPSDVAVVMFRDGCALLQPSKFAISQSGASHSLTISDLVLEDAGQITVEAEGASSAA 1980  
QY 1981 LRVRAPVLFKKLEPQTVESRSSVTLVELTRPKPBLRWRNATAPGNVVEIHAEGA 2040  
Db 1981 LRVRAPVLFKKLEPQTVESRSSVTLVELTRPKPBLRWRNATAPGNVVEIHAEGA 2040  
QY 2041 RHRVLVHNVGFADRGFFCGCEPDDKTQAKLTVENRQVRLVRGLQAVEAREQGATMEVQL 2100  
Db 2041 RHRVLVHNVGFADRGFFCGCEPDDKTQAKLTVENRQVRLVRGLQAVEAREQGATMEVQL 2100  
QY 2101 SHADVGSWTRDGLRFQOGPTCHLAVRGPMTLTLGLRPEDSGIMVFKAEVHTSARLV 2160  
Db 2101 SHADVGSWTRDGLRFQOGPTCHLAVRGPMTLTLGLRPEDSGIMVFKAEVHTSARLV 2160  
QY 2161 VTELVPFSRPLQDVTTEKEKVTLECELSRPNDVRWLKDGVELRAGKTMATAOQACR 2220  
Db 2161 VTELVPFSRPLQDVTTEKEKVTLECELSRPNDVRWLKDGVELRAGKTMATAOQACR 2220  
QY 2221 SLTIYRCEFAFQGVVCDHAQSSASVKVGQRTYTLIYRRVLAEDAGEIQFVAENAESR 2280  
Db 2221 SLTIYRCEFAFQGVVCDHAQSSASVKVGQRTYTLIYRRVLAEDAGEIQFVAENAESR 2280  
QY 2281 ACLVSKELPVTLRPLRKIAMKHGVLCECOVSPASQVRFKGSQELQPGPKYELVSD 2340  
Db 2281 ACLVSKELPVTLRPLRKIAMKHGVLCECOVSPASQVRFKGSQELQPGPKYELVSD 2340  
QY 2341 GLYRKLIISDVHAEDEDYTCADGVKTSQAQFFVEEQSIITVIRGLQDVTVMEPAPAMFEC 2400  
Db 2341 GLYRKLIISDVHAEDEDYTCADGVKTSQAQFFVEEQSIITVIRGLQDVTVMEPAPAMFEC 2400  
QY 2401 ETSIPSVRPPKLLGKTVLQAGNVGLBOGTVHRLMLRRTCTMTGPVHFTVCKSRSSA 2460  
Db 2401 ETSIPSVRPPKLLGKTVLQAGNVGLBOGTVHRLMLRRTCTMTGPVHFTVCKSRSSA 2460  
QY 2461 RLIVSDIPVLTLPKTPKTRGELQSVLSCDFRPAPKAVQYKDDTPLSPSEKFKMSLEG 2520  
Db 2461 RLIVSDIPVLTLPKTPKTRGELQSVLSCDFRPAPKAVQYKDDTPLSPSEKFKMSLEG 2520  
QY 2521 QMAELRIILRLMPADAGVYRCQAGSAHSSTEVTVAREVTVTGPIQDABATEEGWASFSCE 2580  
Db 2521 QMAELRIILRLMPADAGVYRCQAGSAHSSTEVTVAREVTVTGPIQDABATEEGWASFSCE 2580  
QY 2581 LSHDEEVEWSLNGMPLVNDSFHETSHKGRHTLVLSIQRADAGIVRASSLKVSTARS 2640  
Db 2581 LSHDEEVEWSLNGMPLVNDSFHETSHKGRHTLVLSIQRADAGIVRASSLKVSTARS 2640  
QY 2641 EVRVKPVVFLKALDLSAEERGTALQCEVSDPEAHVVRKDGVOQLGSPDKYDLHTAGT 2700  
Db 2641 EVRVKPVVFLKALDLSAEERGTALQCEVSDPEAHVVRKDGVOQLGSPDKYDLHTAGT 2700  
QY 2701 RGLVVDHVSPEADAGLYTCHVGSSETRARVRVHDLHVGITKBLKTMVELEGESCSFECVLS 2760  
Db 2701 RGLVVDHVSPEADAGLYTCHVGSSETRARVRVHDLHVGITKBLKTMVELEGESCSFECVLS 2760  
QY 2761 HESASDPAMWTVGGKTVCSSSRFQATROGRKYLIVREAAAPSDAGEVVVFSVRGLITSKASL 2820  
Db 2761 HESASDPAMWTVGGKTVCSSSRFQATROGRKYLIVREAAAPSDAGEVVVFSVRGLITSKASL 2820  
QY 2821 IYRERPAALIKPLEDQWVAPGEDVELRCELSSAGTPVHMLKDKRAIKRSQKYDVVCEGTM 2880

Db 2821 IYRERPAALIKPLEDQWVAPGEDVELRCELSSAGTPVHMLKDKRAIKRSQKYDVVCEGTM 2880  
QY 2881 AMLVTRGASLKDAGEYTCBEVASKSTASLHVEKANCFTTEELTNLOVEKGTAVTCKTKE 2940  
Db 2881 AMLVTRGASLKDAGEYTCBEVASKSTASLHVEKANCFTTEELTNLOVEKGTAVTCKTKE 2940  
QY 2941 HPAATVTVWKGLELRASGHKQPSQEGTLRLTITISALEKADSDTYTCIDIGQASQAQLIV 3000  
Db 2941 HPAATVTVWKGLELRASGHKQPSQEGTLRLTITISALEKADSDTYTCIDIGQASQAQLIV 3000  
QY 3001 QGRRVHIIBDLBEDVDVQEGSSATFRCPISIPANYEPVHMFPLDKTPLHANELNBIIDAPQGY 3060  
Db 3001 QGRRVHIIBDLBEDVDVQEGSSATFRCPISIPANYEPVHMFPLDKTPLHANELNBIIDAPQGY 3060  
QY 3061 HVLTLRQLALKDSGTITYPFAGDORASAAALRVTEKPSVFSRELTDAITIEGEDTLVCEFS 3120  
Db 3061 HVLTLRQLALKDSGTITYPFAGDORASAAALRVTEKPSVFSRELTDAITIEGEDTLVCEFS 3120  
QY 3121 TCDIPWCMTKDGKTLRGSARCOLSHEGHRAQLLITGATLQDSGRYKCEAGGACSSSIVRV 3180  
Db 3121 TCDIPWCMTKDGKTLRGSARCOLSHEGHRAQLLITGATLQDSGRYKCEAGGACSSSIVRV 3180  
QY 3181 HARVPVFOALKDLEVLLEGGAATLRVLSVAAPVKWCYGNVLPDGYSLRQEGAMLE 3240  
Db 3181 HARVPVFOALKDLEVLLEGGAATLRVLSVAAPVKWCYGNVLPDGYSLRQEGAMLE 3240  
QY 3241 LVVRNLRPDOSGRYSCSGDQTTSATLTVTALPAQFIGKLRNKEATEGATATLRCELSKT 3300  
Db 3241 LVVRNLRPDOSGRYSCSGDQTTSATLTVTALPAQFIGKLRNKEATEGATATLRCELSKT 3300  
QY 3301 APVWRKGSSETLRDGDYCLRDQGANCAELOIRGLAMVDAAEYSCVCGEERTSASLTIRM 3360  
Db 3301 APVWRKGSSETLRDGDYCLRDQGANCAELOIRGLAMVDAAEYSCVCGEERTSASLTIRM 3360  
QY 3361 PAHFIGRLRHQSIIEGATATLRCELSKAAPVWRKGRSLRDGRHSLRDQAVCELOIC 3420  
Db 3361 PAHFIGRLRHQSIIEGATATLRCELSKAAPVWRKGRSLRDGRHSLRDQAVCELOIC 3420  
QY 3421 GLAVADAGYSVCVCGEERTSATLTVKALPAKTEGLRNEEAVEGATAMLWCELSKVAPVE 3480  
Db 3421 GLAVADAGYSVCVCGEERTSATLTVKALPAKTEGLRNEEAVEGATAMLWCELSKVAPVE 3480  
QY 3481 WRKGPENLRDGRYTLRQEGTRCELOICGLANADAGEYLCVCGOERTSATLIRALPARF 3540  
Db 3481 WRKGPENLRDGRYTLRQEGTRCELOICGLANADAGEYLCVCGOERTSATLIRALPARF 3540  
QY 3541 IEDVKNQEARREGATAVLOCELNSAAPVWRKGSSETLRDGDYSLRDQGTKCELOIRGLAM 3600  
Db 3541 IEDVKNQEARREGATAVLOCELNSAAPVWRKGSSETLRDGDYSLRDQGTKCELOIRGLAM 3600  
QY 3601 ADTGEYSVCVCGOERTSATLTVRALPIKFTTEGLRNEEATEGATAVLRCELSKMAPVEMWK 3660  
Db 3601 ADTGEYSVCVCGOERTSATLTVRALPIKFTTEGLRNEEATEGATAVLRCELSKMAPVEMWK 3660  
QY 3661 HETLRDGRHSLRDQGANCAELOIRGLVAEDAGEYLCVCGKERTSAMLTVRAMPSPFIEGL 3720  
Db 3661 HETLRDGRHSLRDQGANCAELOIRGLVAEDAGEYLCVCGKERTSAMLTVRAMPSPFIEGL 3720  
QY 3721 RNEEATEGATATLWCELSKAAPVWRKKGHTETLRDGRHSLRDQGRSCELOIRGLAVVDAG 3780  
Db 3721 RNEEATEGATATLWCELSKAAPVWRKKGHTETLRDGRHSLRDQGRSCELOIRGLAVVDAG 3780  
QY 3781 EYSVCVCGOERTSATLTVRALPARFIEDVKNQEARREGATAVLOCELSKAAPVWRKGSSETL 3840  
Db 3781 EYSVCVCGOERTSATLTVRALPARFIEDVKNQEARREGATAVLOCELSKAAPVWRKGSSETL 3840  
QY 3841 RGGDRYSLRDQGTTRCELOIRGLSVADTGEYSVCVCGOERTSATLTVRAPQVFPREPQSLQ 3900  
Db 3841 RGGDRYSLRDQGTTRCELOIRGLSVADTGEYSVCVCGOERTSATLTVRAPQVFPREPQSLQ 3900  
QY 3901 AEEGSGTATLWCELSSEPTATVWMSKGLQLOQANGRAEPRLQGTAEVLQDREDGEYTT 3960

Db	3901	ABEGSTATLQCELSIPTAVVWSXGGLQLOQANGRRPRLOCTAELVLQDLQREDTGEYT	3960	QY	5041	EPEEPADQTYREDDBHFCIRFEALTEARQAVTRFQEMFATLIGIVEIKLVEQGPRRVEM	5100
QY	3961	CTCSQATSAITLVTAAAPVRELRLOHQEVDEGGTAHLCCELSRAGASVEMRKGSLQFP	4020	Db	5041	EPEEPADQTYREDDBHFCIRFEALTEARQAVTRFQEMFATLIGIVEIKLVEQGPRRVEM	5100
Db	3961	CTCSQATSAITLVTAAAPVRELRLOHQEVDEGGTAHLCCELSRAGASVEMRKGSLQFP	4020	QY	5101	CLSKSTPAPVPPPEPLPSLLTSDAAPVFLTELQNOQEVODGYVPSFDCVVTGQPMPSVRWF	5160
QY	4021	CAKQMVODGAAAEELLVRGVQEBDAGDYTCDTGHTQSMASLSVVRPPKPKTRILQSLQE	4080	Db	5101	CLSKSTPAPVPPPEPLPSLLTSDAAPVFLTELQNOQEVODGYVPSFDCVVTGQPMPSVRWF	5160
Db	4021	CAKQMVODGAAAEELLVRGVQEBDAGDYTCDTGHTQSMASLSVVRPPKPKTRILQSLQE	4080	QY	5161	KDGKLEEDDDHYMINEDQGGHQLIITAVVPADMGVYRCLAEANGSVSSTKAEIRVDLTS	5220
QY	4081	TGDIARLCCQLSDAESGAVVQWLKXGVVHLAGPKYEMRSOCATRELLIHOLEAKDTGEYA	4140	Db	5161	KDGKLEEDDDHYMINEDQGGHQLIITAVVPADMGVYRCLAEANGSVSSTKAEIRVDLTS	5220
Db	4081	TGDIARLCCQLSDAESGAVVQWLKXGVVHLAGPKYEMRSOCATRELLIHOLEAKDTGEYA	4140	QY	5221	TDYDTAADATSSSYFSAQGYLSREQEGTSTTDEGQLPOVYBELRDLQVAPGTRLAKF	5280
QY	4141	CVTGQKTAASLRVTEPEVTIVRGVDAEVTADDEVEFSCEVSRAGATGVQWCLQGLPLQ	4200	Db	5221	TDYDTAADATSSSYFSAQGYLSREQEGTSTTDEGQLPOVYBELRDLQVAPGTRLAKF	5280
Db	4141	CVTGQKTAASLRVTEPEVTIVRGVDAEVTADDEVEFSCEVSRAGATGVQWCLQGLPLQ	4200	QY	5281	OLKVKGYPAPLXWPKDQGPLTASAHIRMTCKILHTLEIISVTRDSGOVAAVINSAMG	5340
QY	4201	SNEVTEAVRGRHTLRLKGVTPEDAGTYSFHLGNHASSAQLTVRAPEVTILEPLQDVQ	4260	Db	5281	OLKVKGYPAPLXWPKDQGPLTASAHIRMTCKILHTLEIISVTRDSGOVAAVINSAMG	5340
Db	4201	SNEVTEAVRGRHTLRLKGVTPEDAGTYSFHLGNHASSAQLTVRAPEVTILEPLQDVQ	4260	QY	5341	AAVSSARLLVRGPDPPEKPSADVHEQLVPPRMLEFPPKVKKGSSITFSVKVEGPRVP	5400
QY	4261	LSEGQDASFOCLRSRASQEARWALGGVPLQANEMNDITVEQGTJHLTLHLKVTLEDAGT	4320	Db	5341	AAVSSARLLVRGPDPPEKPSADVHEQLVPPRMLEFPPKVKKGSSITFSVKVEGPRVP	5400
Db	4261	LSEGQDASFOCLRSRASQEARWALGGVPLQANEMNDITVEQGTJHLTLHLKVTLEDAGT	4320	QY	5401	TVHMLREBAERGVLWIGPDTFGYTVASSAQOHSVLVDVGRQHGTYCTIASNAAGQALC	5460
QY	4321	VSFHVGTCSSEAOIKVTAQNTVVRGLENVBALEGGREALFECOLSOPEVAATHLLDDEPV	4380	Db	5401	TVHMLREBAERGVLWIGPDTFGYTVASSAQOHSVLVDVGRQHGTYCTIASNAAGQALC	5460
Db	4321	VSFHVGTCSSEAOIKVTAQNTVVRGLENVBALEGGREALFECOLSOPEVAATHLLDDEPV	4380	QY	5461	SASLHVSGLPKVEGEKVKKEALISTFLOQTQAIASQOGLTASPADLGGQKKEPLAAKE	5520
QY	4381	RTSENAEVVFPENGRLHLLKNLRPODSRVITPLAGDMVTSAPLTVRGWRLEILPLKN	4440	Db	5461	SASLHVSGLPKVEGEKVKKEALISTFLOQTQAIASQOGLTASPADLGGQKKEPLAAKE	5520
Db	4381	RTSENAEVVFPENGRLHLLKNLRPODSRVITPLAGDMVTSAPLTVRGWRLEILPLKN	4440	QY	5521	ALGHLSLAEVTEEPLOKLTSQITEMVSAKITQAKLVPGGDSDEDESKTSPASPHGRSR	5580
QY	4441	AAVRAGAQAARTCTLSEAVPVGEASWYINGAAVOPDSDSDTWTADGSHQALLRSQPHH	4500	Db	5521	ALGHLSLAEVTEEPLOKLTSQITEMVSAKITQAKLVPGGDSDEDESKTSPASPHGRSR	5580
Db	4441	AAVRAGAQAARTCTLSEAVPVGEASWYINGAAVOPDSDSDTWTADGSHQALLRSQPHH	4500	QY	5581	PSSSTQESSSESSEDGDARGEIFDIYVVTADYPLGAEQDAITLREGQVVEVLDAAHPLRW	5640
QY	4501	AGEVTFACRDVAVASARITLGLPDPPEDAEVVAHSHSTVTLSSWAAPMSSDGGGLCGYRVE	4560	Db	5581	PSSSTQESSSESSEDGDARGEIFDIYVVTADYPLGAEQDAITLREGQVVEVLDAAHPLRW	5640
Db	4501	AGEVTFACRDVAVASARITLGLPDPPEDAEVVAHSHSTVTLSSWAAPMSSDGGGLCGYRVE	4560	QY	5641	LVRTPTKSSPSRQGWNSPAYLDRRLKLSPEWGAABAPPEPCEBAVSEDEYKARLSVIOE	5700
QY	4561	VKEGATGOWRLCHELVPGEPCVVDGLAPGETYFRVAAVGPVAGBEPVHLPTQVRLAEP	4620	Db	5641	LVRTPTKSSPSRQGWNSPAYLDRRLKLSPEWGAABAPPEPCEBAVSEDEYKARLSVIOE	5700
Db	4561	VKEGATGOWRLCHELVPGEPCVVDGLAPGETYFRVAAVGPVAGBEPVHLPTQVRLAEP	4620	QY	5701	LLSSSOAVFVELQFLOSHHLOHLERCCHPVIYVAGQKAVIPRNVRDTCGRHSSFLQELQ	5760
QY	4621	KPVPPQSPAPESROVAAAGEDVLSLEVVAAAGEVVIHKGMERIQPGGRFVVSQGRQOML	4680	Db	5701	LLSSSOAVFVELQFLOSHHLOHLERCCHPVIYVAGQKAVIPRNVRDTCGRHSSFLQELQ	5760
Db	4621	KPVPPQSPAPESROVAAAGEDVLSLEVVAAAGEVVIHKGMERIQPGGRFVVSQGRQOML	4680	QY	5761	CDTDDVAMCFTKQAAFEQYLEFLVGRVQAESVVVSTAIQEFYKKAEEALLAGDPSQP	5820
QY	4681	VIKGFTAEDQEXYHCGLAQSGICPAATFOVALSPASVDEAPQSLPPEAAQSGDLLELLW	4740	Db	5761	CDTDDVAMCFTKQAAFEQYLEFLVGRVQAESVVVSTAIQEFYKKAEEALLAGDPSQP	5820
Db	4681	VIKGFTAEDQEXYHCGLAQSGICPAATFOVALSPASVDEAPQSLPPEAAQSGDLLELLW	4740	QY	5821	PPPLQHYLEQPVREVQRYQALLKELIRNKARNONCALLBQAVVVSALPQRAENKLHV	5880
QY	4741	BALARKMRMSBEPTLDSISELPEEDGSRQRLPOAEAEVADPLSEGYSYTADELARTGDAL	4800	Db	5821	PPPLQHYLEQPVREVQRYQALLKELIRNKARNONCALLBQAVVVSALPQRAENKLHV	5880
Db	4741	BALARKMRMSBEPTLDSISELPEEDGSRQRLPOAEAEVADPLSEGYSYTADELARTGDAL	4800	QY	5881	SIMENYPTLEALGEPTRQGHFIIVMEGAPGARMKPNHNRHVFLFRNHLVTCRPRDSRT	5940
QY	4801	SHTSSDDESAGTSLVTYLLKAGRPCTSPASKVGAPAPSPVKPQOQQPPLAARVPLG	4860	Db	5881	SIMENYPTLEALGEPTRQGHFIIVMEGAPGARMKPNHNRHVFLFRNHLVTCRPRDSRT	5940
Db	4801	SHTSSDDESAGTSLVTYLLKAGRPCTSPASKVGAPAPSPVKPQOQQPPLAARVPLG	4860	QY	5941	DTVSVFRNMKLSISIDNDQVGGDDRAFEVWQREDSVRKYLLOASTAITIKSSWKEIC	6000
QY	4861	DLSTKDLGDPMSDKAAVKIQAAFKGYKVRKEMKQOEGPMFSHTFGDTEAQVGDALRECV	4920	Db	5941	DTVSVFRNMKLSISIDNDQVGGDDRAFEVWQREDSVRKYLLOASTAITIKSSWKEIC	6000
Db	4861	DLSTKDLGDPMSDKAAVKIQAAFKGYKVRKEMKQOEGPMFSHTFGDTEAQVGDALRECV	4920	QY	6001	GIQORLALPVVRPDPFEBELADCTAELGETVKLACRVGTGTPKPIVSWYKDGKAVQVDPFH	6060
QY	4921	VASKADYRARWLKDGVELTDGRHHHIDQLGDGTCSLIAGLDRADAGCYTCQVSNKFGOV	4980	Db	6001	GIQORLALPVVRPDPFEBELADCTAELGETVKLACRVGTGTPKPIVSWYKDGKAVQVDPFH	6060
Db	4921	VASKADYRARWLKDGVELTDGRHHHIDQLGDGTCSLIAGLDRADAGCYTCQVSNKFGOV	4980	QY	6061	ILIEDPDGSCALIILDSITGVDSGOYMCFAAAGNCSTLKGILLVOVPRFVKNKVRASPFV	6120
QY	4981	THSACVVVSGSESAESSSGGELDDAFRRARLRLHRLFRTKSPAEVSDDEELFLSADGPA	5040	Db	6061	ILIEDPDGSCALIILDSITGVDSGOYMCFAAAGNCSTLKGILLVOVPRFVKNKVRASPFV	6120
Db	4981	THSACVVVSGSESAESSSGGELDDAFRRARLRLHRLFRTKSPAEVSDDEELFLSADGPA	5040				

QY 6121 EGEDAQFTCTIEGAPYPOIRYKDGALLTTGNKQTLSEPRSGLLVLVIRAASKEDLGLY 6180  
DB 6121 EGEDAQFTCTIEGAPYPOIRYKDGALLTTGNKQTLSEPRSGLLVLVIRAASKEDLGLY 6180  
QY 6181 ECELVNLGARSASAEALRIQSPMLQAQOCHREQLVAAVEDTTLERADQEVTVSLKELLG 6240  
DB 6181 ECELVNLGARSASAEALRIQSPMLQAQOCHREQLVAAVEDTTLERADQEVTVSLKELLG 6240  
QY 6241 PKAGPSTGDLTGCPGPRGAPALCEGTGQSPVVTGTSEAPVAPPRVPOPLLHHEGPQOEPE 6300  
DB 6241 PKAGPSTGDLTGCPGPRGAPALCEGTGQSPVVTGTSEAPVAPPRVPOPLLHHEGPQOEPE 6300  
QY 6301 A1ARAQEWTVPIRMGAAMPAGAGTCELLWDVHSHVRETTORTVYQAIIDTHARPPSMQ 6360  
DB 6301 A1ARAQEWTVPIRMGAAMPAGAGTCELLWDVHSHVRETTORTVYQAIIDTHARPPSMQ 6360  
QY 6361 VTIEDVQAQTGGAFAEAIIEGDPQSPVTWKDQSVQLVDSRLSQOQEGTYSVLVSHVA 6420  
DB 6361 VTIEDVQAQTGGAFAEAIIEGDPQSPVTWKDQSVQLVDSRLSQOQEGTYSVLVSHVA 6420  
QY 6421 SKDAGVYTCIAQNTGGQVLCFAELVLVGGDNEPSEKQSHRKLHSFYEKKEEIGRGVFG 6480  
DB 6421 SKDAGVYTCIAQNTGGQVLCFAELVLVGGDNEPSEKQSHRKLHSFYEKKEEIGRGVFG 6480  
QY 6481 FVKVQHKGNKILCAAKFIPLRSRTAQAYRERDILAAALSHPLVTGLDQFETRKTLILI 6540  
DB 6481 FVKVQHKGNKILCAAKFIPLRSRTAQAYRERDILAAALSHPLVTGLDQFETRKTLILI 6540  
QY 6541 LELCSSBELLDRLYRKGWVTAEVKVIYIQQLVEGLHVLHSHGVHLHDIKPSNIMLVHHPAR 6600  
DB 6541 LELCSSBELLDRLYRKGWVTAEVKVIYIQQLVEGLHVLHSHGVHLHDIKPSNIMLVHHPAR 6600  
QY 6601 EDIKICDPGFAQNTTPAELQSVGSPFEVSPPEIIQQNPVSEASDIWANGVVISVLSLTC 6660  
DB 6601 EDIKICDPGFAQNTTPAELQSVGSPFEVSPPEIIQQNPVSEASDIWANGVVISVLSLTC 6660  
QY 6661 SPFAGESDRATLLNVLEGRVSWSPMAHLSEDAKOFIKATLQAPQAPSAACLSHPW 6720  
DB 6661 SPFAGESDRATLLNVLEGRVSWSPMAHLSEDAKOFIKATLQAPQAPSAACLSHPW 6720  
QY 6721 FLKSNPABEAFINTKQJFLARSWORSLSWYSKSLVMSISPELLRGPPDSLSGLVAR 6780  
DB 6721 FLKSNPABEAFINTKQJFLARSWORSLSWYSKSLVMSISPELLRGPPDSLSGLVAR 6780  
QY 6781 HLCDRTGSS 6840  
DB 6781 HLCDRTGSS 6840  
QY 6841 RSTAPAPAPGEGAGPAAQGCVPVRSVIRSLFYHQAGSEPHGALAPGSRHRPARRH 6900  
DB 6841 RSTAPAPAPGEGAGPAAQGCVPVRSVIRSLFYHQAGSEPHGALAPGSRHRPARRH 6900  
QY 6901 LLKGGYIAGALPGLREPLMEHRVLEEEAAREEQATLLAKAPSFETALRLPASGTHLAPGH 6960  
DB 6901 LLKGGYIAGALPGLREPLMEHRVLEEEAAREEQATLLAKAPSFETALRLPASGTHLAPGH 6960  
QY 6961 SHSLEHDSFSTPRSSSEACGAQRLLPSAPSGAPIRDMGHPPQGSKQLPSTGCGHPTGAQPE 7020  
DB 6961 SHSLEHDSFSTPRSSSEACGAQRLLPSAPSGAPIRDMGHPPQGSKQLPSTGCGHPTGAQPE 7020  
QY 7021 RPSDPSMGQAPFCHPQGSAPQGCSPHPAVAPCPGSPPGCSKEAPLVPSSPFLGQ 7080  
DB 7021 RPSDPSMGQAPFCHPQGSAPQGCSPHPAVAPCPGSPPGCSKEAPLVPSSPFLGQ 7080  
QY 7081 PQAPPAPAKAPPLDKNKPGDILSLGRPKGPCSSPGSASQASSSSQVSSLRVSSQVGT 7140  
DB 7081 PQAPPAPAKAPPLDKNKPGDILSLGRPKGPCSSPGSASQASSSSQVSSLRVSSQVGT 7140  
QY 7141 EPGPSLDAGWTOEAEDLSDSTPTLQRPQEQATWRKFSLGRGGYAGVAGYGTAFAGGDA 7200  
DB 7141 EPGPSLDAGWTOEAEDLSDSTPTLQRPQEQATWRKFSLGRGGYAGVAGYGTAFAGGDA 7200  
QY 7201 GGMLQGGPMWARIANAVSQSEEEQBEARAEASQSEEQARAEASPLPQVSARPPVEVGRA 7260

DB 7201 GGMLQGGPMWARIANAVSQSEEEQBEARAEASQSEEQARAEASPLPQVSARPPVEVGRA 7260  
QY 7261 PTRSSPEPTWEDIGQVSLVQIRDLSGDAEAAADTISLDISEVDPAYLNLSDLYDKYLPF 7320  
DB 7261 PTRSSPEPTWEDIGQVSLVQIRDLSGDAEAAADTISLDISEVDPAYLNLSDLYDKYLPF 7320  
QY 7321 EPMIFKVPKSAQPPPPSPMAEEELAEPEPTWMPGELGPHAGLEITEESDNDVALLAE 7380  
DB 7321 EPMIFKVPKSAQPPPPSPMAEEELAEPEPTWMPGELGPHAGLEITEESDNDVALLAE 7380  
QY 7381 AAVGRKRKWSPPSRSLFHPGSHLPLDDEPAELGLRERVKASVEHISRLIKGRPEGLEKEG 7440  
DB 7381 AAVGRKRKWSPPSRSLFHPGSHLPLDDEPAELGLRERVKASVEHISRLIKGRPEGLEKEG 7440  
QY 7441 PRKKEGLASFLRSLGKSWDRAPTFRLRELSDETVVLGQSVTLACQVSAQPAQAATWSKDG 7500  
DB 7441 PRKKEGLASFLRSLGKSWDRAPTFRLRELSDETVVLGQSVTLACQVSAQPAQAATWSKDG 7500  
QY 7501 APLESSRVLISATLKNFOLLTILVVABDLGVTCVSNALGTVTITGVLKKAERPSS 7560  
DB 7501 APLESSRVLISATLKNFOLLTILVVABDLGVTCVSNALGTVTITGVLKKAERPSS 7560  
QY 7561 PCPDIGEVADGVLVWKPVEYGPVTYIYQCSLEGGSWTTLASDIFDCCYLTSKLSRG 7620  
DB 7561 PCPDIGEVADGVLVWKPVEYGPVTYIYQCSLEGGSWTTLASDIFDCCYLTSKLSRG 7620  
QY 7621 TYTFRACVSKAGMGPYSPSEQVLLGGPSHLASEEESQGRSAQPLPSTKTFAFOTQIOR 7680  
DB 7621 TYTFRACVSKAGMGPYSPSEQVLLGGPSHLASEEESQGRSAQPLPSTKTFAFOTQIOR 7680  
QY 7681 GRFVVVRQCEKASGALAAKIIPYHKDKTAVLREYEALKGLRHPHLAQLHAAVLSPRH 7740  
DB 7681 GRFVVVRQCEKASGALAAKIIPYHKDKTAVLREYEALKGLRHPHLAQLHAAVLSPRH 7740  
QY 7741 LVLILELCSGPPELLPCLAEASYSSEVKDYLMQWLSATQYLNQHLHLDRSENMIIT 7800  
DB 7741 LVLILELCSGPPELLPCLAEASYSSEVKDYLMQWLSATQYLNQHLHLDRSENMIIT 7800  
QY 7801 EYNLLKVVDLGNALQSQKVLPSDKFDYLETMAPELLEGOGAVPQTDIWAIGVTAFTM 7860  
DB 7801 EYNLLKVVDLGNALQSQKVLPSDKFDYLETMAPELLEGOGAVPQTDIWAIGVTAFTM 7860  
QY 7861 LSAEYVPSSEGARDLQRLKGLVRLSRCYAGLSGGAVAFRLSTLCAQPWGRPCASSCLQ 7920  
DB 7861 LSAEYVPSSEGARDLQRLKGLVRLSRCYAGLSGGAVAFRLSTLCAQPWGRPCASSCLQ 7920  
QY 7921 CPWLTEEGPACSPAPVPTPTARLVFVNRKRALLVKRHNLAQVR 7968  
DB 7921 CPWLTEEGPACSPAPVPTPTARLVFVNRKRALLVKRHNLAQVR 7968  
RESULT 2  
ID ABG76186  
XX ID ABG76186 standard; protein; 2630 AA.  
AC ABG76186;  
XX  
DT 09-MAY-2003 (first entry)  
XX  
DE Human serine/threonine or protein kinase 59079.  
XX Human; enzyme; serine/threonine kinase; protein kinase; 59079;  
XX cardiovascular disease; heart failure; myocardial infarction;  
XX blood vessel disorder; atherosclerosis; Kaposi's sarcoma; immunogen;  
XX blood platelet disorder; thrombocytopenia; leukaemia; Hodgkin's disease;  
XX haemolytic anaemia; cellular proliferative disorder; cancer;  
XX protein kinase disorder; autoimmune disorder; diabetes mellitus;  
XX psoriasis; inflammatory bowel disease; rheumatoid arthritis;  
XX multiple sclerosis.  
OS Homo sapiens.  
XX

PN	US2002168742-A1.	QY	5579	SRPSSSIQESSSESDGDARGEI	FDIYVVVTADYPLGAEQDAITLREGQYVEVLDAAHPL	5638
XX		DB	241	SRPSSSIQESSSESDGDARGEI	FDIYVVVTADYPLGAEQDAITLREGQYVEVLDAAHPL	300
PD	14-NOV-2002.					
XX						
PF	15-FEB-2002; 2002US-00077130.	QY	5639	RWLVRTKPTKSSPSRQGWSPAYLDRRLKLSPEWGAAEAPPEPGEAVSDEYKARLSSVI	5698	
XX		DB	301	RWLVRTKPTKSSPSRQGWSPAYLDRRLKLSPEWGAAEAPPEPGEAVSDEYKARLSSVI	360	
XX						
PR	15-FEB-2001; 2001US-0269201P.	QY	5699	QELLSSEQAFFVEELQFLQSHHLQHLERCPHPVPIAVAGQKAVIFRNVRDIGRHFSSFLQEL	5758	
XX		DB	361	QELLSSEQAFFVEELQFLQSHHLQHLERCPHPVPIAVAGQKAVIFRNVRDIGRHFSSFLQEL	420	
PA	(MILL-) MILLENNIUM PHARM INC.					
PI	Kapeller-Libermann R, Acton SL;					
XX		QY	5759	QOCDDDDVAMCFIKNOAAFEQYLEFLVGRVQAVSVVYSTAIQEFYKKAEEALLAGDPS	5818	
DR	WPI; 2003-298729/29.	DB	421	QOCDDDDVAMCFIKNOAAFEQYLEFLVGRVQAVSVVYSTAIQEFYKKAEEALLAGDPS	480	
XX	N-PSDB; ABX11641.					
PT	Novel isolated human protein kinase, designated 59079 or 12599	QY	5819	QPPPPLOHYLEQPVVERVQRYQALLKELIRKARNRQNCALLEQAYAVVSALPQRAENKL	5878	
PT	polypeptide, useful as diagnostic and therapeutic agents for preventing	DB	481	QPPPPLOHYLEQPVVERVQRYQALLKELIRKARNRQNCALLEQAYAVVSALPQRAENKL	540	
PT	cardiovascular diseases, proliferative disorders, and protein kinase					
PT	disorders.					
XX						
PS	Claim 8; Page 48-54; 119pp; English.	QY	5879	HVSLMENYPGTLEALGEPHFIQGHFIVWEGAPGARMKGNHNVFLFRNHILVICPRRDS	5938	
XX		DB	541	HVSLMENYPGTLEALGEPHFIQGHFIVWEGAPGARMKGNHNVFLFRNHILVICPRRDS	600	
CC	The invention relates to an isolated human serine/threonine or protein	QY	5939	RTDTVSVFRNMKLSIDLDNDQVEGDDRAFEVWQEREDSVKYLQARTAIKSSWVKE	5998	
CC	kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule	DB	601	RTDTVSVFRNMKLSIDLDNDQVEGDDRAFEVWQEREDSVKYLQARTAIKSSWVKE	660	
CC	comprising at least 85% identity to the nucleic acids appearing as					
CC	ABX11641 and ABX11642 or their complement, a naturally occurring variant	QY	5999	ICGICQELALPWRRPDPFEELADCTAELGETVKLACRVGTGPKPVISWYKDGKAVQVDP	6058	
CC	of the kinases or their fragments. Also included are a non-human host	DB	661	ICGICQELALPWRRPDPFEELADCTAELGETVKLACRVGTGPKPVISWYKDGKAVQVDP	720	
CC	cell containing the nucleic acids, an antibody specific for the proteins,	QY	6059	HHILIEDPGSCALILDSLTGVDGQYMCFAASAGNCSTLGKILVQVPPFRVNVKVRASP	6118	
CC	identifying a compound which binds to the kinase (by contacting the	DB	721	HHILIEDPGSCALILDSLTGVDGQYMCFAASAGNCSTLGKILVQVPPFRVNVKVRASP	780	
CC	kinase or a cell expressing the kinase with a test compound) and					
CC	determining whether the kinase binds to the test compound) and modulating	QY	6119	FVEGEDAQFTCTIEGAPYQIRWYKDGALLTGKFNKFTLSEPRSGLLVIVIRAAKEDLG	6178	
CC	the activity of kinase using the identified compound. The kinases and	DB	781	FVEGEDAQFTCTIEGAPYQIRWYKDGALLTGKFNKFTLSEPRSGLLVIVIRAAKEDLG	840	
CC	their encoding nucleic acids are useful as diagnostic and therapeutic	QY	6179	LYECELNVRLGSARASAEIRIQSPMLQAEQCHREQLVAIVEDTLERADQVTSVLKRL	6238	
CC	agents for preventing a disease or condition associated with an aberrant	DB	841	LYECELNVRLGSARASAEIRIQSPMLQAEQCHREQLVAIVEDTLERADQVTSVLKRL	900	
CC	or unwanted 59079 or 12599 activity in a subject, including	QY	6239	LGPKAPGSTGDLTGPGCPRGAPALQETGSPVVTGTSEAPVPRVPQPLHGGEPQE	6298	
CC	cardiovascular diseases such as heart failure, and myocardial infarction;	DB	901	LGPKAPGSTGDLTGPGCPRGAPALQETGSPVVTGTSEAPVPRVPQPLHGGEPQE	960	
CC	disorders involving blood vessels such as atherosclerosis, and Kaposi's	QY	6299	PEAIAQAQEWTPVIRMEGAAMPAGAGTBELLWDVHSHVVRETTQRTYTYQAIDHTARPPS	6358	
CC	sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia,	DB	961	PEAIAQAQEWTPVIRMEGAAMPAGAGTBELLWDVHSHVVRETTQRTYTYQAIDHTARPPS	1020	
CC	Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders	QY	6359	MOVTTIEDVQAGTGTGAQFEALIEGDPQSPVTVYKDSVOLVDSTRLSQOQEGTTYSLVRH	6418	
CC	such as cancer; and protein kinase disorders such as autoimmune	DB	1021	MOVTTIEDVQAGTGTGAQFEALIEGDPQSPVTVYKDSVOLVDSTRLSQOQEGTTYSLVRH	1080	
CC	disorders, diabetes mellitus, psoriasis, inflammatory bowel disease,	QY	6419	VASKDAGVYVTLCAQNTGGQVLCKAELLVLGGDNEPDSEKQSHRRKLHSFYEVKEIEGRGV	6478	
CC	rheumatoid arthritis, and multiple sclerosis (many examples of diseases	DB	1081	VASKDAGVYVTLCAQNTGGQVLCKAELLVLGGDNEPDSEKQSHRRKLHSFYEVKEIEGRGV	1140	
CC	and disorders are included in the specification). The kinases, their	QY	6479	FGFVKRVQHKGNKILCAAKFIPLRSRTRAQAYRERDILAAASHPLVTGLDQFETKTLI	6538	
CC	encoding nucleic acids and antibodies are useful in screening assays,	DB	1141	FGFVKRVQHKGNKILCAAKFIPLRSRTRAQAYRERDILAAASHPLVTGLDQFETKTLI	1200	
CC	detection assays (e.g. forensic biology), and predictive medicine (e.g.	QY	6539	LILELCSSEBLLDRLYRKGVVTEAEVKVYIQQLVEGLHYLHSHGVHLHLDIKPSNLMVHP	6598	
CC	diagnostic assays, prognostic assays, and monitoring clinical trials and	DB	1201	LILELCSSEBLLDRLYRKGVVTEAEVKVYIQQLVEGLHYLHSHGVHLHLDIKPSNLMVHP	1260	
CC	pharmacogenomics). The kinases and their encoding nucleic acids are	QY	6599	AREDIKICDFGFAQNTIPAELOFQYQSGPFPVSPPEITQQNPVSEASDIWAMGVTSYLSLT	6658	
CC	useful as query sequences to perform a search against public databases to	DB	1261	AREDIKICDFGFAQNTIPAELOFQYQSGPFPVSPPEITQQNPVSEASDIWAMGVTSYLSLT	1320	
CC	identify other family members or related sequences. The present sequence	QY	6659	CSSPAGESDRAITLNVLEGRVSNSSPMAHLSEDAKFIKATLQAPQAPPSAAQCULSH	6718	
CC	represents kinase 59079					
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX						
XX		</				

Sequence 2630 AA;

Query Match 33.2%; Score 13710; DB 6; Length 2630;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2630; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	5339	MGAAYSSARLLVRGDPDEPEKPSADVHEQLVPPRMLERFTPKVKVKGSSITFSVKVEGRP	5398
DB	1	MGAAYSSARLLVRGDPDEPEKPSADVHEQLVPPRMLERFTPKVKVKGSSITFSVKVEGRP	60
QY	5399	VPTVHVLREEAERGVLWIGDTPGYTVASSAQHSLVLLDVGRQHOGTYTCIASNAAGQA	5458
DB	61	VPTVHVLREEAERGVLWIGDTPGYTVASSAQHSLVLLDVGRQHOGTYTCIASNAAGQA	120
QY	5459	LCSASLVSGLPKVEQEKKYKEALISTFLOGTTQAIQAQGLTASPADLGGQKEPLAA	5518
DB	121	LCSASLVSGLPKVEQEKKYKEALISTFLOGTTQAIQAQGLTASPADLGGQKEPLAA	180
QY	5519	KEALGHLSLAERGTEEFQKLQTSITQVMSAKITQAKLVQPGGDSDDSKTSPASPRHGR	5578
DB	181	KEALGHLSLAERGTEEFQKLQTSITQVMSAKITQAKLVQPGGDSDDSKTSPASPRHGR	240

1321 CSSPAGESDRATLNLVLEGRVSSPMAAHLSDAKDFIKATLQRAFQAPRPAQAQCLSH 1380  
6719 PWF1KSPMAEBAHINTKOLKFLIARSRWORSLSYKSIILVMSIPELLRGPDPSPSLGV 6778  
1381 PWF1KSPMAEBAHINTKOLKFLIARSRWORSLSYKSIILVMSIPELLRGPDPSPSLGV 1440  
6779 ARHLCRTDGSSSSSSSSDNEAPARAKSLPPSPVTHSPILLHPRGFLRPSASLPEAEAE 6938  
1441 ARHLCRTDGSSSSSSSSDNEAPARAKSLPPSPVTHSPILLHPRGFLRPSASLPEAEAE 1500  
6939 SERTEAPAPASPEGAGPAPACVPRHSVIRSLFYHQGESPEHGALAPGRRHPARR 6998  
1501 SERTEAPAPASPEGAGPAPACVPRHSVIRSLFYHQGESPEHGALAPGRRHPARR 1560  
6899 RHLLKGGYIAGALPGLREPLMEHRLVLEBEAREEQATILLAKAPSFETALRLPASGTHLAP 6958  
1561 RHLLKGGYIAGALPGLREPLMEHRLVLEBEAREEQATILLAKAPSFETALRLPASGTHLAP 1620  
6959 GHSHSLHDSPTPRPSEACGEAQRLLPSAPSGGAPIRDMGHPOGSKQLPSTGCHPGTAQ 7018  
1621 GHSHSLHDSPTPRPSEACGEAQRLLPSAPSGGAPIRDMGHPOGSKQLPSTGCHPGTAQ 1680  
7019 PERSPSPWQAPAFCHPKQGSAPQSGSPHAPVAPCPGSPGPPGCKEAPLVPSSPFL 7078  
1681 PERSPSPWQAPAFCHPKQGSAPQSGSPHAPVAPCPGSPGPPGCKEAPLVPSSPFL 1740  
7079 GQOAPAPAKASPLPSKMGPGDISLPGPKPGCSPPGASQASSQVSSLRVSSQV 7138  
1741 GQOAPAPAKASPLPSKMGPGDISLPGPKPGCSPPGASQASSQVSSLRVSSQV 1800  
7139 GTEPGPSLDAGWTQAEADLSDTPTLQRPQEQATMRKFSLGGGGYAGVAGTGFAGG 7198  
1801 GTEPGPSLDAGWTQAEADLSDTPTLQRPQEQATMRKFSLGGGGYAGVAGTGFAGG 1860  
7199 DAGMLGQGPWARIANAVSQQESEEAEAEQSEEEQAEAESELPQVSRARPVEVG 7258  
1861 DAGMLGQGPWARIANAVSQQESEEAEAEQSEEEQAEAESELPQVSRARPVEVG 1920  
7259 RAPTRSSPEPTPWEIDIGVSLVQIRDLSDGAEADTTSLDISEVDPRAYLNSLDYDKYL 7318  
1921 RAPTRSSPEPTPWEIDIGVSLVQIRDLSDGAEADTTSLDISEVDPRAYLNSLDYDKYL 1980  
7319 PFEMIFRKVKQSPQPEPPMAEELAEFPETWMPGELPHAGLEITESESDVDALL 7378  
1981 PFEMIFRKVKQSPQPEPPMAEELAEFPETWMPGELPHAGLEITESESDVDALL 2040  
7379 ABAVGRKRWSSPERSLFHPGRHLPLDEPAELGLRERVKASVEHISRLKGRPEGLEK 7438  
2041 ABAVGRKRWSSPERSLFHPGRHLPLDEPAELGLRERVKASVEHISRLKGRPEGLEK 2100  
7439 EGPRKPKGLASFRSLGKSDRAPTFRLRLSDETVLQSVTLACQVSAQPAQAATWSK 7498  
2101 EGPRKPKGLASFRSLGKSDRAPTFRLRLSDETVLQSVTLACQVSAQPAQAATWSK 2160  
7499 DGAPLESSRVLISATLKNFOLLITLVVAEDLGVYTCVSNALGTVTITGVLEKABRPS 7558  
2161 DGAPLESSRVLISATLKNFOLLITLVVAEDLGVYTCVSNALGTVTITGVLEKABRPS 2220  
7559 SSPCPDICEVADGVLLVWKPVESGPTYIIVQCSLEGGSWTTLASDIFDCCYLTKLSR 7618  
2221 SSPCPDICEVADGVLLVWKPVESGPTYIIVQCSLEGGSWTTLASDIFDCCYLTKLSR 2280  
7619 GGTYTRFACVSKAGMGVSPSEQVLLGGPSHLASBESQGRSAQPLPSTKTFAFQTQI 7678  
2281 GGTYTRFACVSKAGMGVSPSEQVLLGGPSHLASBESQGRSAQPLPSTKTFAFQTQI 2340  
7679 QGRFSVVRQCKEASGALAAKIIPYHPDKTAVLREYKALGLRHPLAQLHAAYLSP 7738  
2341 QGRFSVVRQCKEASGALAAKIIPYHPDKTAVLREYKALGLRHPLAQLHAAYLSP 2400  
7739 RHLVLILELCSGPELLPCLAEASVSESEVKDYLWQMLSATQYLNHQLHLDLRSENI 7798

2401 RHLVLILELCSGPELLPCLAEASVSESEVKDYLWQMLSATQYLNHQLHLDLRSENI 2460  
7799 ITEYNLLKVVDLGNAGSLSQEKVLPSPDKFDYLETMAPELLEGGQAVPQTDIWAIGVTAF 7858  
2461 ITEYNLLKVVDLGNAGSLSQEKVLPSPDKFDYLETMAPELLEGGQAVPQTDIWAIGVTAF 2520  
7859 IMLSAYFPVSSSGARDLQRLKGLVRLSRVRCVAGLSGGAVAFRLSTICAOPHRCASSC 7918  
2521 IMLSAYFPVSSSGARDLQRLKGLVRLSRVRCVAGLSGGAVAFRLSTICAOPHRCASSC 2580  
7919 LQCPWLTEGPACSRPAPVTFPTAFLRVFVRNREKRALLYKRNHLAQVR 7968  
2581 LQCPWLTEGPACSRPAPVTFPTAFLRVFVRNREKRALLYKRNHLAQVR 2630

RESULT 3  
AAB30569  
ID AAB30569 standard; protein; 2596 AA.  
XX  
AC AAB30569;  
XX  
DT 19-MAR-2001 (first entry)  
XX  
DE A splice variant of a signal transduction polypeptide.  
XX  
KW Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;  
congestive heart failure; dilated congestive cardiomyopathy;  
hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;  
mitral valve disease; aortic valve disease; tricuspid valve disease;  
myocardial infarction; cardiac arrhythmia; arteriosclerosis;  
atherosclerosis; cardiac tumour; microbial infection; splice variant.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FT Domain 325..504 Location/Qualifiers  
FT Domain /note= "guanine nucleotide exchange factor domain"  
FT Domain 1094..1351  
FT Domain /note= "kinase domain"  
FT Domain 2301..2553  
FT Domain /note= "kinase domain"  
XX  
FN WO200063381-A1.  
XX  
XX  
PD 26-OCT-2000.  
XX  
PF 11-APR-2000; 2000WO-US009488.  
XX  
PR 16-APR-1999; 99US-0129553P.  
XX  
PA (SCIO-) SCIOS INC.  
XX  
PI Zeng W, Stanton L, Kong H;  
XX  
XX  
DR WPI; 2001-007013/01.  
DR N-PSDB; AAC62287.  
XX  
XX  
PT Novel h19G5 polypeptides capable of regulating signal transduction and  
exhibiting kinase activity useful for identifying antibodies to treat  
cardiac diseases, and additional mediators of signal transduction.  
XX  
PS Claim 1; Page 68-74; 81pp; English.  
XX  
CC The present sequence represents a splice variant of human in signal  
transduction polypeptide. The polypeptide is designated H19G5. The  
protein is capable of regulating signal transduction and exhibits kinase  
activity. The H19G5 transcript is expressed in the heart. H19G5  
polypeptides and polynucleotides are useful for preventing or treating a  
cardiac disease, such as congestive heart failure, dilated congestive  
cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,  
mitral valve disease, aortic valve disease or tricuspid valve disease,  
angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,  
arterial or renovascular hypertension, arteriosclerosis, atherosclerosis

CC	and cardiac tumours in humans. The polypeptide is also useful for
CC	detecting the expression of a protein capable of regulating signal
CC	transduction or the expression of a protein capable of acting as a donor
CC	or acceptor molecule of a phosphate group. The monoclonal antibodies can
CC	be used as probes for detecting discrete antigens expressed by tissue or
CC	cell samples, and therefore used in humans for localization and
CC	monitoring of microbial infection
xx	
SQ	Sequence 2596 AA;
	Query Match 32.8%; Score 13528; DB 4; Length 2596;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 2595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	5373 MLERFTPKKVKKSSITFSVKEGRPVPTVHVLREEAERGVLWIGDTPGYTVASSAQQH 5432
DB	1 MLERFTPKKVKKSSITFSVKEGRPVPTVHVLREEAERGVLWIGDTPGYTVASSAQQH 60
QY	5433 SLVLLDVGRQHQGTCTCIASNAAGALCSALHVSGLPKVEEQKVKREALISIFLOQTQ 5492
DB	61 SLVLLDVGRQHQGTCTCIASNAAGALCSALHVSGLPKVEEQKVKREALISIFLOQTQ 120
QY	5493 AISAQGLTASFDLGGORKEEPLAAKEALCHLSLAEVGTEEFLOKLTSTITWVSAKIT 5552
DB	121 AISAQGLTASFDLGGORKEEPLAAKEALCHLSLAEVGTEEFLOKLTSTITWVSAKIT 180
QY	5553 QAKLVQPGDSEDESKTPSASPRHGRSRPSSSIOESSESEGDARGEIIFIYVVTADYL 5612
DB	181 QAKLVQPGDSEDESKTPSASPRHGRSRPSSSIOESSESEGDARGEIIFIYVVTADYL 240
QY	5613 PLGAEQDAITUREQYVEVLDAHLPLWLVRTKTSRSGQWNSPAYLDRLKLSPEW 5672
DB	241 PLGAEQDAITUREQYVEVLDAHLPLWLVRTKTSRSGQWNSPAYLDRLKLSPEW 300
QY	5673 GAAEAPFPGEAVSEDEYKARLSSVIOELLSSQAFVEELQFLQSHLQHLERCPHVPIA 5732
DB	301 GAAEAPFPGEAVSEDEYKARLSSVIOELLSSQAFVEELQFLQSHLQHLERCPHVPIA 360
QY	5733 VAGQAVIFRNVRDIGRFHSSFLQELQCCDTDDVAMCFIKNOAAFEQYLEFVLGVQAE 5792
DB	361 VAGQAVIFRNVRDIGRFHSSFLQELQCCDTDDVAMCFIKNOAAFEQYLEFVLGVQAE 420
QY	5793 SVVYSTAIQEFYKKAEBEALLAGDPSPPPPQLCHYLEQVVERVQRYQALLKELIRNKAR 5852
DB	421 SVVYSTAIQEFYKKAEBEALLAGDPSPPPPQLCHYLEQVVERVQRYQALLKELIRNKAR 480
QY	5853 NRQNCALLEQYAVVSALPQAEKHLVSLMENYPTGLEALGPIRQGHFIWEGAPGAR 5912
DB	481 NRQNCALLEQYAVVSALPQAEKHLVSLMENYPTGLEALGPIRQGHFIWEGAPGAR 540
QY	5913 MPWKGNHRHFLFRNLHVICPKPRDSDRTDVSIVFRNMKLSIDLNDQVGGDDRAFEVW 5972
DB	541 MPWKGNHRHFLFRNLHVICPKPRDSDRTDVSIVFRNMKLSIDLNDQVGGDDRAFEVW 600
QY	5973 QEREDSVRKYLQARTAIKSSWVKEICGIQORLALPWRPDPFESEELADCTAELGETVK 6032
DB	601 QEREDSVRKYLQARTAIKSSWVKEICGIQORLALPWRPDPFESEELADCTAELGETVK 660
QY	6033 LACRVGTGTPKRVISWYKDGKAVQVDPHHILIEDPDGSCALILDSLTGVDSGQWCFPAASA 6092
DB	661 LACRVGTGTPKRVISWYKDGKAVQVDPHHILIEDPDGSCALILDSLTGVDSGQWCFPAASA 720
QY	6093 AGNCSTLGLKILVQVPPFRVNVKVRASPEVEGEDAQFTCTIECAPYQIRWYKDGALLTTGN 6152
DB	721 AGNCSTLGLKILVQVPPFRVNVKVRASPEVEGEDAQFTCTIECAPYQIRWYKDGALLTTGN 780
QY	6153 KFQTLSPRGLVILVTRAASKEDLGLYECELVNRGLSARASAEIRIQSPMLQAOQCHR 6212
DB	781 KFQTLSPRGLVILVTRAASKEDLGLYECELVNRGLSARASAEIRIQSPMLQAOQCHR 840
QY	6213 EQLVAAVEDTTLERADQVTSVLKRLIGPKAPGPGSTGDLTGPGFCPRGAPALQETGQPP 6272
DB	841 EQLVAAVEDTTLERADQVTSVLKRLIGPKAPGPGSTGDLTGPGFCPRGAPALQETGQPP 900

QY	6273 VTGTSEAPVPRVQPOLHHEGPEQPEAIARAQEWTPVIRMEGAAMPAGTIGELLWDVH 6332
DB	901 VTGTSEAPVPRVQPOLHHEGPEQPEAIARAQEWTPVIRMEGAAMPAGTIGELLWDVH 960
QY	6333 SHVRETTQRTYTYQAIDHTHTARPPSMQVTIEDVQAQTGGTAQFEAIIIEGDPQPSVTWYK 6392
DB	961 SHVRETTQRTYTYQAIDHTHTARPPSMQVTIEDVQAQTGGTAQFEAIIIEGDPQPSVTWYK 1020
QY	6393 DSVQLVDSSTELSQOQEGTYSILVRHVASKDAGVYVTCIAQNTGGVLCABELLVGGDNE 6452
DB	1021 DSVQLVDSSTELSQOQEGTYSILVRHVASKDAGVYVTCIAQNTGGVLCABELLVGGDNE 1080
QY	6453 PDSEKQSHRRKHLHSFYEVKEEIGRGVFGFVKRVQHGKNKILCAAKFIPLRSRTRAQAYRE 6512
DB	1081 PDSEKQSHRRKHLHSFYEVKEEIGRGVFGFVKRVQHGKNKILCAAKFIPLRSRTRAQAYRE 1140
QY	6513 RDILAAALSHPLVTGLLDQFETRKTILILILELCSSEELLDRLYRKGVVTEAEVKYIQQV 6572
DB	1141 RDILAAALSHPLVTGLLDQFETRKTILILILELCSSEELLDRLYRKGVVTEAEVKYIQQV 1200
QY	6573 EGLHYLHSHGVHLHLDIKPSNIMLVHPAREDIKICDFGPAQNITPAELQFSQYGSPEFVSP 6632
DB	1201 EGLHYLHSHGVHLHLDIKPSNIMLVHPAREDIKICDFGPAQNITPAELQFSQYGSPEFVSP 1260
QY	6633 EIIQONPVSEASDIWAMGVI SYLSITCSPSPAGESDRAITLLNVLEGRVSWSSPVAHLSE 6692
DB	1261 EIIQONPVSEASDIWAMGVI SYLSITCSPSPAGESDRAITLLNVLEGRVSWSSPVAHLSE 1320
QY	6693 DAKDIFKATLQAPQAPRASAQCLSHPMFLKSMAPAEAEHFINTKQLFLARSWQBSLM 6752
DB	1321 DAKDIFKATLQAPQAPRASAQCLSHPMFLKSMAPAEAEHFINTKQLFLARSWQBSLM 1380
QY	6753 SYKSLVNVRSIPELLRGPPDSPSLGVARHLCRDTGGSSSSSSSDNELAPPARAKSLPPS 6812
DB	1381 SYKSLVNVRSIPELLRGPPDSPSLGVARHLCRDTGGSSSSSSSDNELAPPARAKSLPPS 1440
QY	6813 PVTHSPHLHPGFLRPSASLPEEAESRSTEARPPASPEGAPPAAGCVCPHSHVIRS 6872
DB	1441 PVTHSPHLHPGFLRPSASLPEEAESRSTEARPPASPEGAPPAAGCVCPHSHVIRS 1500
QY	6873 LFYHQAGSEPHGALAPGSRHRRHLLKGGYIAGALPGLREPLMEHVRLESEAAARE 6932
DB	1501 LFYHQAGSEPHGALAPGSRHRRHLLKGGYIAGALPGLREPLMEHVRLESEAAARE 1560
QY	6933 QATLLAKAPSETALRLPASGTHLAPGSHSHLEHDSSTPRPSSEACGEAQLRPSAPSGG 6992
DB	1561 QATLLAKAPSETALRLPASGTHLAPGSHSHLEHDSSTPRPSSEACGEAQLRPSAPSGG 1620
QY	6993 APIRDMGHPQSKQLPSTGGHPGTAQPERPSPDPSPGWQPAFFCHPKGSGAPQEGSCSPHA 7052
DB	1621 APIRDMGHPQSKQLPSTGGHPGTAQPERPSPDPSPGWQPAFFCHPKGSGAPQEGSCSPHA 1680
QY	7053 VAPCPGSPFGSCKEAPLVPSSPFLQOPQAPPAKASPPDLKMGPGDISLPGRKXPG 7112
DB	1681 VAPCPGSPFGSCKEAPLVPSSPFLQOPQAPPAKASPPDLKMGPGDISLPGRKXPG 1740
QY	7113 PCSFPGSASQASSQVSSLRVGSQVGTPEGPSLDAEGWTQAEADLSDSTPTLQRPQEA 7172
DB	1741 PCSFPGSASQASSQVSSLRVGSQVGTPEGPSLDAEGWTQAEADLSDSTPTLQRPQEA 1800
QY	7173 TMRKPSLGGRGYAGVAGYGTFAFGGAGMGLGGPMWARIAMAVSSEEEEBEAEARAS 7232
DB	1801 TMRKPSLGGRGYAGVAGYGTFAFGGAGMGLGGPMWARIAMAVSSEEEEBEAEARAS 1860
QY	7233 QSEEQEAEARASPLPQVSARVPVPGRAPTRHSSPEPTWEDIGQVSLVQIRDLSGDAEA 7292
DB	1861 QSEEQEAEARASPLPQVSARVPVPGRAPTRHSSPEPTWEDIGQVSLVQIRDLSGDAEA 1920
QY	7293 DTISLIDISEVDPAYLNLSLDYDIKYLPEFMI FRKVPKSAQPEPPSPMAEEELAEFPPEPT 7352
DB	1921 DTISLIDISEVDPAYLNLSLDYDIKYLPEFMI FRKVPKSAQPEPPSPMAEEELAEFPPEPT 1980

QY 7353 WPMFGLPGHAGLEITESESDVALLAAEAVGKRKWSPPSRSLFHPFGRLPLDEPAEL 7412  
Db 1981 WPMFGLPGHAGLEITESESDVALLAAEAVGKRKWSPPSRSLFHPFGRLPLDEPAEL 2040  
QY 7413 GLREVKASVEHISRLKGPGLKLEKGGPRKPGGLASPELSGLKSWDRAPTFLRELSDE 7472  
Db 2041 GLREVKASVEHISRLKGPGLKLEKGGPRKPGGLASPELSGLKSWDRAPTFLRELSDE 2100  
QY 7473 TVVLGQSVTLACQVSAQPAQAATWSKDGAPLESSRVLISATLKNFOLLTILVVVAEDLG 7532  
Db 2101 TVVLGQSVTLACQVSAQPAQAATWSKDGAPLESSRVLISATLKNFOLLTILVVVAEDLG 2160  
QY 7533 VYTCVSNALGTVTTGVLKRAERPSSPCPDIGEVYADGVLLVWKPVESYGPVTYIVOC 7592  
Db 2161 VYTCVSNALGTVTTGVLKRAERPSSPCPDIGEVYADGVLLVWKPVESYGPVTYIVOC 2220  
QY 7593 SLEGGSWTTLASDIFDCCYLTSLKSGTTFRTACVSKAGMGPYSSPSQVLLGPFSLH 7652  
Db 2221 SLEGGSWTTLASDIFDCCYLTSLKSGTTFRTACVSKAGMGPYSSPSQVLLGPFSLH 2280  
QY 7653 ASESESQGRSAQPLPSKTFAQTQIORGRFSVVRQCWEKASGRALAAKIIPYHPKDKTA 7712  
Db 2281 ASESESQGRSAQPLPSKTFAQTQIORGRFSVVRQCWEKASGRALAAKIIPYHPKDKTA 2340  
QY 7713 VLREYALKGLRHPLHQAOLHAAVLSPRHLVLIILELCSGPPELLCLAERASYSESEVKDYL 7772  
Db 2341 VLREYALKGLRHPLHQAOLHAAVLSPRHLVLIILELCSGPPELLCLAERASYSESEVKDYL 2400  
QY 7773 WQMLSATQYLHNOHILHDLRSENMIITEYNLLKVVDLGNASQSKVLPSDKFKDYLE 7832  
Db 2401 WQMLSATQYLHNOHILHDLRSENMIITEYNLLKVVDLGNASQSKVLPSDKFKDYLE 2460  
QY 7833 TMAPLEGGQAVPQTDIWAIGVTFIMLSAEYVPSSEGARDLQGRKGLVRLSRCYAG 7892  
Db 2461 TMAPLEGGQAVPQTDIWAIGVTFIMLSAEYVPSSEGARDLQGRKGLVRLSRCYAG 2520  
QY 7893 LSGGAVAFRLSTLCAQWGRPCASSCLQCPWLTEGPACSRPAPVTFTTARLXVFVNRE 7952  
Db 2521 LSGGAVAFRLSTLCAQWGRPCASSCLQCPWLTEGPACSRPAPVTFTTARLXVFVNRE 2580  
QY 7953 KRRALLYKRHNLAQVR 7968  
Db 2591 KRRALLYKRHNLAQVR 2596  
RESULT 4  
ABP70084  
XX ID ABP70084 standard; protein; 4591 AA.  
XX AC ABP70084;  
XX DT 27-JAN-2003 (first entry)  
XX DE Human NOV13a.  
XX KW Human; anti-HIV; cytostatic; antidiabetic; antiasthmatic; cachexia; AIDS;  
XX KW antiinflammatory; cardiant; haemostatic; neuroprotective; anorectic;  
XX KW neutropic; immunosuppressive; osteopathic; antiparkinsonian; cancer;  
XX KW antifertility; cerebroprotective; gene therapy; NOVX; NOV; fertility;  
XX KW metabolic disorder; diabetes; obesity; infectious disease; anorexia;  
XX KW neurodegenerative diseases; Alzheimer's disease; Parkinson's disease;  
XX KW immune disorder; haematopoietic disorder; cardiovascular disorder;  
XX KW bronchial asthma; dyslipidemia; metabolic disturbance; neurogenesis;  
XX KW metabolic syndrome X; wasting disorder; cell differentiation;  
XX KW cell proliferation; haematopoiesis; wound healing; angiogenesis.  
XX OS Homo sapiens.  
XX XX WO200272771-A2.  
XX PN 19-SEP-2002.  
XX PD 08-MAR-2002; 2002WO-US007288.  
XX PF

XX 08-MAR-2001; 2001US-0274101P.  
PR 08-MAR-2001; 2001US-0274194P.  
PR 08-MAR-2001; 2001US-0274281P.  
PR 08-MAR-2001; 2001US-0274322P.  
PR 09-MAR-2001; 2001US-0274849P.  
PR 12-MAR-2001; 2001US-0275235P.  
PR 13-MAR-2001; 2001US-0275578P.  
PR 13-MAR-2001; 2001US-0275579P.  
PR 13-MAR-2001; 2001US-0275601P.  
PR 14-MAR-2001; 2001US-0276000P.  
PR 16-MAR-2001; 2001US-0276776P.  
PR 19-MAR-2001; 2001US-0276994P.  
PR 20-MAR-2001; 2001US-0277239P.  
PR 20-MAR-2001; 2001US-0277321P.  
PR 20-MAR-2001; 2001US-0277327P.  
PR 20-MAR-2001; 2001US-0277338P.  
PR 21-MAR-2001; 2001US-0277791P.  
PR 22-MAR-2001; 2001US-0277833P.  
PR 23-MAR-2001; 2001US-0278152P.  
PR 26-MAR-2001; 2001US-0278894P.  
PR 27-MAR-2001; 2001US-0278999P.  
PR 27-MAR-2001; 2001US-0279036P.  
PR 28-MAR-2001; 2001US-0279344P.  
PR 30-MAR-2001; 2001US-0279995P.  
PR 30-MAR-2001; 2001US-0280233P.  
PR 02-APR-2001; 2001US-0280802P.  
PR 02-APR-2001; 2001US-0280822P.  
PR 02-APR-2001; 2001US-0280900P.  
PR 04-APR-2001; 2001US-0281194P.  
PR 13-APR-2001; 2001US-0283675P.  
PR 30-APR-2001; 2001US-0287424P.  
PR 02-MAY-2001; 2001US-0288066P.  
PR 03-MAY-2001; 2001US-0288342P.  
PR 03-MAY-2001; 2001US-0288528P.  
PR 15-MAY-2001; 2001US-0291190P.  
PR 16-MAY-2001; 2001US-0291099P.  
PR 16-MAY-2001; 2001US-0291240P.  
PR 30-MAY-2001; 2001US-0294485P.  
PR 31-MAY-2001; 2001US-0294889P.  
PR 31-MAY-2001; 2001US-0294899P.  
PR 18-JUN-2001; 2001US-0299027P.  
PR 19-JUN-2001; 2001US-0299303P.  
PR 19-JUN-2001; 2001US-0299310P.  
PR 10-JUL-2001; 2001US-0304354P.  
PR 31-JUL-2001; 2001US-0309198P.  
PR 16-AUG-2001; 2001US-0312903P.  
PR 10-SEP-2001; 2001US-0318462P.  
PR 12-SEP-2001; 2001US-0318770P.  
PR 27-SEP-2001; 2001US-0325430P.  
PR 27-SEP-2001; 2001US-0325681P.  
PR 18-OCT-2001; 2001US-0330380P.  
PR 31-OCT-2001; 2001US-0335301P.  
PR 14-NOV-2001; 2001US-0332172P.  
PR 14-NOV-2001; 2001US-0332271P.  
PR 14-NOV-2001; 2001US-0332272P.  
PR 14-NOV-2001; 2001US-0333184P.  
PR 21-NOV-2001; 2001US-0333272P.  
PR 03-DEC-2001; 2001US-0332094P.  
PR 03-DEC-2001; 2001US-0337426P.  
PR 04-DEC-2001; 2001US-0338092P.  
PR 03-JAN-2002; 2001US-0337185P.  
PR 08-MAR-2002; 2002US-0345705P.  
PR 08-MAR-2002; 2002US-00093463.  
(CURA-) CURAGEN CORP.  
XX Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;  
XX PI Boldog FI, Li L, Zerhusen BD, Tchernev VT, Gangoli EA, Vernet CAM;  
XX PI Pena CEA, Burgess CE, Liu X, Spytek KA, Gorman L, Spaderna SK;  
XX PI Voss EZ, Malyankar UM, Anderson DW, Patturajan M, Miller CE;  
XX PI Taupier RJ, Padigar M, Shenoy SG, Kekuda R, Gusev VY, Fochart PF;  
XX PI Zhong M;



XX WPI; 2002-732824/79.  
 DR N-PSDB; ABV99362.  
 XX  
 PT New NOVX polypeptides and polynucleotides, useful for preventing,  
 PT diagnosing or treating NOVX-associated disorders e.g. diabetes, cancer,  
 PT Alzheimer's disease, dyslipidemias, obesity, immune or hematopoietic  
 PT disorders, and asthma.  
 XX  
 PS Claim 1; Page 137-138; 619pp; English.  
 XX  
 CC The present invention relates to new isolated proteins (NOVX) and their  
 CC coding sequences (ABV99327-ABV99595 and ABP70049-ABP70149), where X is  
 CC any number from 1 to 48. The NOVX proteins and coding sequences are  
 CC useful in the manufacture of a medicament for treating a syndrome  
 CC associated with a human disease, preferably a NOVX-associated disorder.  
 CC The NOVX coding sequences and proteins are useful for treating, diabetes,  
 CC preventing or diagnosing diseases such as metabolic disorders, diabetes,  
 CC obesity, infectious disease, anorexia, cancer-associated cachexia,  
 CC cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's  
 CC disease, immune disorders, hematopoietic disorders, cardiovascular  
 CC disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic  
 CC disturbances associated with obesity, metabolic syndrome X or wasting  
 CC disorders associated with chronic diseases or various cancers. The NOVX  
 CC coding sequences and proteins may also be used as targets for the  
 CC identification of small molecules that modulate or inhibit e.g.  
 CC neurogenesis, cell differentiation, cell proliferation, haematopoiesis,  
 CC wound healing and angiogenesis, in gene therapy, in generation of  
 CC antibodies that bind immunospecifically to NOVX substances for use in  
 CC therapeutic or diagnostic methods  
 XX  
 SQ Sequence 4691 AA;

Query Match 25.5%; Score 10519.5; DB 5; Length 4691;  
 Best Local Similarity 51.1%; Pred. No. 0;  
 Matches 2367; Conservative 118; Mismatches 342; Indels 1807; Gaps 66;

QY 3135 LRGSARCOLSHEGHAQLLITGATLQDSGRYKCEAGGACSSSIVRVHARPVRFQALKDL 3194  
 DB 541 LRGSARCOLSHEGHAQLLITGATLQDSGRYKCEAGGACSSSIVRVHARPVRFQALKDL 600  
 QY 3195 EYLEGAATLRCLVSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVRLNRPDQSGRY 3254  
 DB 601 EYLEGAATLRCLVSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVRLNRPDQSGRY 660  
 QY 3255 SCSEFGDQTSATLTVTALPAQFIGLRNKEATEGATATLRCELSKTAAPVWRKGSSETLRD 3314  
 DB 661 SCSEFGDQTSATLTVTALPAQFIGLRNKEATEGATATLRCELSKTAAPVWRKGSSETLRD 720  
 QY 3315 GDYRCLRDGAMCELIQIRGLAMVDAAEYSCVCGEERTSASLTIRPMAHFTGLRHHQESI 3374  
 DB 721 GDYRCLRDGAMCELIQIRGLAMVDAAEYSCVCGEERTSASLTIRPMAHFTGLRHHQESI 780  
 QY 3375 EGATATLRCELSKTAAPVWRKGSSETLRDGRHSRLRQDGAVALCELOICGLAVADAGEYSCVC 3434  
 DB 781 EGATATLRCELSKTAAPVWRKGSSETLRDGRHSRLRQDGAVALCELOICGLAVADAGEYSCVC 840  
 QY 3435 GEERTSATLTVKALPAKFTGELRNEEAVEGATAMLCELS KVAAPVWRKGPENLRDGRY 3494  
 DB 841 GEERTSATLTVKALPAKFTGELRNEEAVEGATAMLCELS KVAAPVWRKGPENLRDGRY 900  
 QY 3495 ILRQEGTRCELIQICGLAMADAGEYLCVCGQERTSATLTIRALPARFIEDVKNQAREGAT 3554  
 DB 901 ILRQEGTRCELIQICGLAMADAGEYLCVCGQERTSATLTIRALPARFIEDVKNQAREGAT 960  
 QY 3555 AVLOCELNSAAPVWRKGSSETLRDGRYSLRQDGTCKELOIRGLAMADTGEYSCVCCQER 3614  
 DB 961 AVLOCELNSAAPVWRKGSSETLRDGRYSLRQDGTCKELOIRGLAMADTGEYSCVCCQER 1020  
 QY 3615 TSAMLTVRALPIKFTGELRNEEATEGATAVLRCELSKVAAPVWRKGHETLRDGRHSLRQ 3674  
 DB 1021 TSAMLTVRALPIKFTGELRNEEATEGATAVLRCELSKVAAPVWRKGHETLRDGRHSLRQ 1080  
 QY 3675 DGARCELIQIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPKFTGLENEEATEGDTATLW 3734  
 DB 1081 DGARCELIQIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPKFTGLENEEATEGDTATLW 1140  
 QY 3735 CELSKAAPVWRKGHETLRDGRHSLRQDGRSCELOIRGLAVVDAGEYSCVCGQERTSAT 3794  
 DB 1141 CELSKAAPVWRKGHETLRDGRHSLRQDGRSCELOIRGLAVVDAGEYSCVCGQERTSAT 1200  
 QY 3795 LTVRALPARFIEDVKNQAREGATAVLRCELSKVAAPVWRKGSSETLRGGDYSRLRQDGR 3854  
 DB 1201 LTVRALPARFIEDVKNQAREGATAVLRCELSKVAAPVWRKGSSETLRGGDYSRLRQDGR 1260  
 QY 3855 CELQIHGLSVADTGEYSCVCGQERTSATLTVR ----- 3886  
 DB 1261 CELQIHGLSVADTGEYSCVCGQERTSATLTVRALPARFTQDLTKKEASEGATATLQCELS 1320  
 QY 3887 ----- 3886  
 DB 1321 KVAPVWRKGPETLRDGRYSLRQDGTCELOIRGLVADAGEYSCVCGQERTSATLTVR 1380  
 QY 3887 ----- 3886  
 DB 1381 ALPARFTEGLRNEEAMSGATATLQCELSKVAAPVWRKGLEALRDGDYSLRQDGAVALCELO 1440  
 QY 3887 ----- 3886  
 DB 1441 IHGLAMADNGYSCVCGQERTSATLTVRALPARFIEDMRNKNQKATEGATVTLQCKLRKAAP 1500  
 QY 3887 ----- 3886  
 DB 1501 VWRKGPENLRDGRYSLRQDGTCELOIRGLVADAGEYSCVCGQERTSATLTVRALPA 1560  
 QY 3887 ----- 3886  
 DB 1561 RFIEDVRNHEATEGATAVLRCELSKVAAPVWRKGSSETLRDGRYSLRQDGTCELOIRGL 1620  
 QY 3887 ----- 3886



Db	1621	AVEDTGEYLCVCGOERTSATLTVRALPARFIDNMNTQAEAGATATLHCBSKVAVPEWR	1680	Db	2661	VTPEAGTVSFHLGNHASSAQLTVPAPEVTILEPQDVQLR	2701
Qy	3887	-----	3886	Qy	4282	RWALGGVPIQANEMNDITVEQGTLLHLLTHKVTLEDAGTVSFHVCTCSSEALKVT	4337
Db	1681	KGPETLRDGRHSRLQDGRCELQIRGLAVVDAGEYSCVCGOERTSATLTVRALPARFIE	1740	Db	2702	-----GVPLQANEMNDITVEQGTLLHLLTHKVTLEDAGTVSFHVCTCSSEALKVTAVP	2756
Qy	3887	-----	3886	Qy	4338	-----	4337
Db	1741	DVKQAEAGATAVLOCELSKAAPVWRKSGSETLRGGDRYSLRODGRCELOIHLGLSVAD	1800	Db	2757	CLVRGLQNVDPFAGEVATPFCEDGQPSAIAVRDGIHLSMLSLGLGVADSGTVI FRAGPLV	2816
Qy	3887	-----	3886	Qy	4338	-----	4337
Db	1801	TGEYSCVCGOERTSATLTVRALPARFTQDLTKKEASEGATATLOCELSKVAVPEWKKGPE	1860	Db	2817	STAKLLIKDPVVEVVSAMODLAVEBGGSALLCOYSRPVQATWMDREVHTDGRHVIE	2876
Qy	3887	-----	3886	Qy	4338	-----	4337
Db	1861	TLRDGRYSLQDGRCELOIHLGLSVADAGEYSCVCGOERTSATLTVRDCHTLHVMHPY	1920	Db	2877	QOWNVARLTFRPAIPCDSGIYSCAAAGTRVVALLOVQAQNTVVVRGLNVEALEGGEALPE	2936
Qy	3887	-----	3886	Qy	4361	COLSQPEVAHAHTWLLDDEPVRTSENAEVVFPFENGLRHLHLLKNLRPQDSQCRVTFIAGDMV	4420
Db	1921	FQPLGKBPETLIYQIPSPVILFTEGLRNEAMEGATATLOCELSKAAPVWRKGLE	1980	Db	2937	COLSQPEVAHAHTWLLDDEPVRTSENAEVVFPFENGLRHLHLLKNLRPQDSQCRVTFIAGDMV	2996
Qy	3887	-----	3886	Qy	4421	TSAPFLTVR-----GNRLEILEPLKNAAPVAGACAREFTCTLSEAVPVGEASWINGAA	4472
Db	1981	ALRDGKXSLRQDGAVCLOIHLGLAMADNGVYSSIPARFIEDMRNQKATEGATVTLQCKL	2040	Db	2997	TSAPFLTVRGDCAVLVQGWLEILEPLKNAAVRAGACAREFTCTLSEAVPVGEASWINGAA	3056
Qy	3887	-----	3886	Qy	4473	VQPDSDMTVTADGSHQALLRSAPPHHAGEVTFACRDVAVASARLTVLGLPDPPEDAEVV	4532
Db	2041	RKAAPVWRKGPNTLKDGDRYSLQDGTSCBLQIRGLVIADAGYSCICEQERTSATLTV	2100	Db	3057	VQPDSDMTVTADGSHHALLRSAPPHHAGEVTFACRDVAVASARLTVLGLPDPPEDAEVV	3116
Qy	3887	-----	3886	Qy	4533	AKSSHVTLSWAAPMSDGGGLCGYRVEVKEGATQWRLCHELVGPEVCVVDGLARGETY	4592
Db	2101	RALPARFIEDVRNHEATEGATAVLOCELSKAAPVWRKSGSETLRDGRYSLRQDGRCEL	2160	Db	3117	ARSSHTVTLISWAAPMSDGGGLCGYRVEVKEGATQWRLCHELVGPEVCVVDGLARGETY	3176
Qy	3887	-----	3886	Qy	4593	RFRVAAGVPVGAPEVHLPTVRLABPPKPPQPSAPESROVAAGEDVSELEVVVAEAG	4652
Db	2161	QIRGLAVEDTGEYLCVCGOERTSATLTVRALPARFIDNMNTQAEAGATATLHCBSKVAV	2220	Db	3177	RFRVAAGVPVGAPEVHLPTVRLABPPKPPQPSAPESROVAAGEDVSELEVVVAEAG	3235
Qy	3887	-----	3886	Qy	4653	EVTHKGMERIQPGRFEVYVSGRQOMLVIKGFTAEQGEYHCGLAGGSCIPAAATFOVA	4712
Db	2221	PVWRKGPETLRDGRHSRLQDGRCELOIHLGLSVADAGEYSCVCGOERTSATLTVRALP	2280	Db	3236	EVTHKGMERIQPGRFEVYVSGRQOMLVIKGFTAEQGEYHCGLAGGSCIPAAATFOVA	3295
Qy	3887	-----	3886	Qy	4713	LSPASVDEAPQPSLPPEAAQEGDLHLLWEALARKRMSREPTLDSISELPEEDGRQORLP	4772
Db	2281	AKFTKGLRNEATEGATAMLOCELSKVAVPEWRKGPETLRDGRYVNLQDGRCELOIHLG	2340	Db	3296	LSPASVDEAPQPSLPPEAAQEGDLHLLWEALARKRMSREPTLDSISELPEEDGRQORLP	3355
Qy	3887	-----	3886	Qy	4773	QEAEEVAPDLSEGYSTADELARTGDADLSHTSSDDESAGTSLVTLTKKAGRPQTSPLA	4832
Db	2341	LSVADTGEYSCVCGOERTSATLTVKAPQVFPVREPLQSLQAEBSGSTATLQCSLSEPTATV	2400	Db	3356	QEAEEVAPDLSEGYSTADELARTGDADLSHTSSDDESAGTSLVTLTKKAGRPQTSPLA	3415
Qy	3922	WSKGGLOQANGRRPRLOGCTAEVLVLODLOREDTGEYTCGSOATSATLTVTAAPVRF	3981	Qy	4833	SKYGAP-----AASVXKPPQOQOEP	4852
Db	2401	WSKGGLOQANGRRPRLOGCTAEVLVLODLOREDTGEYTCGSOATSATLTVTAAPVRF	2460	Db	3416	SKVSPENLACKERFPTPRAGSLLGVGADPAFPGERSARCTRRCAAPPRESLKRREP	3474
Qy	3982	LRELQHOEVEGGTAHLCELSRAGASVWRKSGSLQLPFCAYQMVQDGAALLVRGVE	4041	Qy	4853	AAVRPPLGLDLSKDLG-----DPSMD-----KAAVKIQ-----AAPK	4884
Db	2461	LRELQHOEVEGGTAHLCELSRAGASVWRKSGSLQLPFCAYQMVQDGAALLVRGVE	2520	Db	3475	ASCLP--GAMEAVELARKLOEBATCILDYFDTPVMTTCGHNFRCACIQLSWEKARGKX	3532
Qy	4042	QEDAGDYTCDDTCHTOSMASLSVRVRPKFKTLQSLQEQETGDIARLCCQLSDAESGAVVQ	4101	Qy	4885	GYKVRK-----EMKQOEGPMFHTFGDTAQVGDALRECVVASKADVBARWLKDG	4935
Db	2521	QEDAGDYTCDDTCHTOSMASLSVRVRPKFKTLQSLQEQETGDIARLCCQLSDAESGAVVQ	4161	Db	3533	GRKRKGSFPCECREMSQORNLN-----RLTKVAEMA-----	3568
Qy	4102	WLKEGVHLAGPKYEMRSOGATRELLIHOLEAKDTGEYACVGTGOKTAAASURVTEPEVTI	4161	Qy	4936	VELTDRGHHHIDQIGDGTCSLLIAGLDRADAGCVTCQVSNK-----FGQVTHSACVVVSGS	4991
Db	2550	-----CGPQVRDAAQAGATRELLIHOLEAKDTGEYACVGTGOKTAAASURVTEPEVTI	2600	Db	3569	-----OQH-----PGLOQD-----LCQEHHEPLKLCQKQCSICVVCRE	3604
Qy	4162	VRGLVDAVTADEDFEFSCEVSRAGATGVOMCLOGLPLQSNVETVAVRDGRIHTLRKLG	4221	Qy	4992	ESEAESGGGLDDAFRAAARRHLRFRKSPAEVSDDEELFLSADEG-----	5038
Db	2601	VRGLVDAVTADEDFEFSCEVSRAGATGVOMCLOGLPLQSNVETVAVRDGRIHTLRKLG	2660	Db	3605	SRE-----HRLHVL-----PAEEAVQGYKJLKEEDMEYLREQITRTGN	3643
Qy	4222	VTPEAGTVSFHLGNHASSAQLTVPAPEVTILEPQDVQLSREGQDASFCRLSRASGQEA	4281	Qy	5039	--PABPEPADWQ--TYREDEHFICIRFEAL-----TEASQAVTRFOEMF	5079
				Db	3644	LQAREOSLAEBWQGVKERRERIVLEPEKVNLYLVEEORLLOALETEEBETASRLRESV	3703

QY 5080 ATL---GIGVEIKLVE-----QGPRR-----VEMCISK-- 5104  
Db 3704 ACLDRGHSLELLLLQLLEERSTQGLQMLQDKMKEPLSRAALLVLIHGMNVLPEFVVSPL 3763  
QY 5105 SPLYLIATKAHTQLGPGTPTDPECPTEPLPISPP-PRSS--TEDVVPDATSAFYLLLYE 5131  
Db 3764 SPLYLIATKAHTQLGPGTPTDPECPTEPLPISPP-PRSS--TEDVVPDATSAFYLLLYE 3820  
QY 5132 LQNEVEQDQYP-----VSFDCVVTQGPMPVSRWFKDGLKLEDDHYMINEQOQ 5180  
Db 3821 SRQRYLGLSSPEGSGFCSDRFVAYPCAV-QQTA-----FSSGR-----HYNEVGMNIT 3868  
QY 5181 GHQLIITAVPADMG-----VYRCIAENSMGVSSTKAELRVDLTSTDYDTAADATESSYF 5236  
Db 3869 GDALWALGVCRDNVSRKDRVPC-K-PENGFWV-----VQLSK-----GTYLYLSTP 3911  
QY 5237 SAQGYLSRE--QEGTSTTDEBQLPOVBEELDLQVAPGTRLAKFKLVKGVYAPRPLY 5293  
Db 3912 SALTPVWMLPEPSSHMGIFLDFEAG-----EVSFYSVSDGSHLHTYSQAT--PPGQLOP 3962  
QY 5294 WFKDQPLTASAHIRMTGKILHTLEIISVTREDSGOVAAVYSNAGAAVSSARLLVRGP 5353  
Db 3963 FFCGLGAP-----KSGQWISVTWAGV-KDLATRTGAVVTPALGA----- 4001  
QY 5354 DEPEKPADSVHEQLVPPRMLERTPKKKKGSSITFSVKVGPVPTVHMLREAEARGV 5413  
Db 4002 -----YAPGATETQS-----PAP-----WSPRAPE-- 4021  
QY 5414 LWIGPDTPTGYVASSAQHSLVLLDVGRHQGTCTCIASNAAGQALCSLSLHVSGLPKVE 5473  
Db 4022 -----PEHPG--VPSLAPRSA-----RACAAAPGYGSPRAA 4051  
QY 5474 BOEKVKEALISTFLQGTQTAISAGLETASPADLGGORKEEPLAAKEALGHLSLAEVGT 5533  
Db 4052 EAARRPADSTAFPL-SVRAMAA-----PDLSTNLQEAATCA----- 4087  
QY 5534 EFLQKLSQITEMYSAKITQAKLOVPGGDSDESKTPSAS--PRHGRSPSSI----- 5585  
Db 4088 ICLDYFTDPVMDCGHNFRCRRCQWQPEARTAPASACPRGTCGPTARLLRWPRW 4147  
QY 5586 QESSESEDGARGEIIDIYVVTADYPLGAE-QDAITL-----REGQVVEVLDAHPL 5638  
Db 4148 RGACTRRRRRR-----VPAHREPLAFCDELLCAACERSGEH----- 4189  
QY 5639 RWL-----VRTKPKSSPSRQ-----WSPAYLDRLK-LSPSE----- 5671  
Db 4190 -WAHRVGRCTRPKTSRPLEAGTMAANETLLSGAKLEKSLHLRQWQDALLFOAQDET 4248  
QY 5672 ---WGAAEAPEFPGEAEVSEYKARLSSVIOELLSQEQAFVEELOFQSHHL-----QHLE 5724  
Db 4249 CVLWQADGGEQORQNVLEFE---RLRLLAEGTAAABAGEEELKQSAHLAELE 4305  
QY 5725 RCPHVPIAVAGQKA 5738  
Db 4306 R--PLPAACAGAAA 4317

RESULT 5

ABP70085

ID ABP70085 standard; protein; 4675 AA.

XX AC

XX AC

XX DT

XX DT 27-JAN-2003 (first entry)

XX DE

XX DE Human NOV13b.

XX Human; anti-HIV; cytostatic; antidiabetic; antiasthmatic; cachexia; AIDS;  
KW antiinflammatory; cardiac; haemostatic; neuroprotective; anorectic;  
KW nootropic; immunosuppressive; osteopathic; antiparkinsonian; cancer;  
KW antifertility; cerebroprotective; gene therapy; NOVX; NOV; fertility;  
KW metabolic disorder; diabetes; obesity; infectious disease; anorexia;

KW neurodegenerative disease; Alzheimer's disease; Parkinson's disease;  
KW immune disorder; haematopoietic disorder; cardiovascular disorder;  
KW bronchial asthma; dyslipidemia; metabolic disturbance; neurogenesis;  
KW metabolic syndrome X; wasting disorder; cell differentiation;  
KW cell proliferation; haematopoiesis; wound healing; angiogenesis.

Homo sapiens.

WO200272771-A2.

XX 19-SEP-2002.

XX 08-MAR-2002; 2002WO-US007288.

XX 08-MAR-2001; 2001US-0274101P.

XX 08-MAR-2001; 2001US-0274194P.

XX 08-MAR-2001; 2001US-0274281P.

XX 08-MAR-2001; 2001US-0274322P.

XX 09-MAR-2001; 2001US-0274849P.

XX 12-MAR-2001; 2001US-0275235P.

XX 13-MAR-2001; 2001US-0275578P.

XX 13-MAR-2001; 2001US-0275799P.

XX 13-MAR-2001; 2001US-0275601P.

XX 16-MAR-2001; 2001US-0276000P.

XX 16-MAR-2001; 2001US-0276766P.

XX 20-MAR-2001; 2001US-0276994P.

XX 20-MAR-2001; 2001US-0277239P.

XX 20-MAR-2001; 2001US-0277321P.

XX 20-MAR-2001; 2001US-0277327P.

XX 20-MAR-2001; 2001US-0277388P.

XX 21-MAR-2001; 2001US-0277791P.

XX 22-MAR-2001; 2001US-0277833P.

XX 23-MAR-2001; 2001US-0278152P.

XX 26-MAR-2001; 2001US-0278894P.

XX 27-MAR-2001; 2001US-0278999P.

XX 27-MAR-2001; 2001US-0279036P.

XX 28-MAR-2001; 2001US-0279344P.

XX 30-MAR-2001; 2001US-0279995P.

XX 30-MAR-2001; 2001US-0280233P.

XX 02-APR-2001; 2001US-0280802P.

XX 02-APR-2001; 2001US-0280822P.

XX 02-APR-2001; 2001US-0280900P.

XX 04-APR-2001; 2001US-0281194P.

XX 13-APR-2001; 2001US-0283675P.

XX 30-APR-2001; 2001US-0287424P.

XX 02-MAY-2001; 2001US-0288066P.

XX 03-MAY-2001; 2001US-0288342P.

XX 15-MAY-2001; 2001US-0288528P.

XX 16-MAY-2001; 2001US-0291099P.

XX 30-MAY-2001; 2001US-0291240P.

XX 31-MAY-2001; 2001US-0294855P.

XX 31-MAY-2001; 2001US-0294899P.

XX 18-JUN-2001; 2001US-0299027P.

XX 19-JUN-2001; 2001US-0299303P.

XX 10-JUL-2001; 2001US-0299310P.

XX 31-JUL-2001; 2001US-0304354P.

XX 16-AUG-2001; 2001US-0309198P.

XX 10-SEP-2001; 2001US-0312903P.

XX 12-SEP-2001; 2001US-0318462P.

XX 27-SEP-2001; 2001US-0318770P.

XX 27-SEP-2001; 2001US-0325430P.

XX 18-OCT-2001; 2001US-0325681P.

XX 31-OCT-2001; 2001US-0330380P.

XX 14-NOV-2001; 2001US-0332172P.

XX 14-NOV-2001; 2001US-0332271P.

XX 14-NOV-2001; 2001US-0332272P.

XX 14-NOV-2001; 2001US-0333184P.

XX 21-NOV-2001; 2001US-0333272P.

XX 03-DEC-2001; 2001US-0332094P.

XX 03-DEC-2001; 2001US-0337426P.

PR	03-DEC-2001; 2001US-0338092P.																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
----	-------------------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

PR	03-DEC-2001; 2001US-0338092P.	QY	2595	MPLYNDSPHEISHKGRRHTLVLSKIQRADAGIVRASSLKVST	SARLEVRKPVVFLKALD	2654		
PR	04-DEC-2001; 2001US-0337185P.	Db	1	MPLYNDSPHEISHKGRRHTLVLSKIQRADAGIVRASSLKVST	SARLEVRKPVVFLKALD	60		
PR	03-JAN-2002; 2002US-0345705P.	QY	2655	DLSAERGTALQCEVSDPEAHVVRKDGVLGSPDKYDFLHTAGT	RGLVHVDPEDAG	2714		
PR	08-MAR-2002; 2002US-00093463.	Db	61	DLSAERGTALQCEVSDPEAHVVRKDGVLGSPDKYDFLHTAGT	RGLVHVDPEDAG	120		
XX	(CURA-) CURAGEN CORP.	QY	2715	LYTCHVGSEETRARVVRHDLRVGKITKLTMTVELEGESCSFECVL	SHESASDPAMTVGG	2774		
XX	Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ; Bolog FL, Li L, Zerhusen BD, Tchernev VT, Gangolli EA, Vernet CAM; Pena CEA, Burgess CE, Liu X, Spytek KA, Gorman L, Spaderna SK; Voss EZ, Malyankar UM, Anderson DW, Patturajan M, Miller CE; Taupier RJ, Padigar M, Shenoy SG, Kekuda R, Gusev VI, Pochart PF; Zhong M;	Db	121	LYTCHVGSEETRARVVRHDLRVGKITKLTMTVELEGESCSFECVL	SHESASDPAMTVGG	180		
XX	WPI; 2002-732824/79.	QY	2775	KTVGSSSRFOATQGRKYILVVR	EAAPSDAGEVVFVSRGLTSKASLIVRERPA	IIKPLE 2834		
XX	N-PSDB; ABV99363.	Db	181	KTVGSSSRFOATQGRKYILVVR	EAAPSDAGEVVFVSRGLTSKASLIVRERPA	IIKPLE 240		
XX	New NOVX polypeptides and polynucleotides, useful for preventing, diagnosing or treating NOVX-associated disorders e.g. diabetes, cancer, Alzheimer's disease, dyslipidemias, obesity, immune or hematopoietic disorders, and asthma.	QY	2835	DQWAPGEDVELRCEL	SRACTPPVHVLKDRKAIRKSQKYDVVCSGT	MAMLVIRGASLKDAG 2894		
XX	Claim 1; Page 142-143; 619pp; English.	Db	241	DQWAPGEDVELRCEL	SRACTPPVHVLKDRKAIRKSQKYDVVCSGT	MAMLVIRGASLKDAG 300		
XX	The present invention relates to new isolated proteins (NOVX) and their coding sequences (ABV99337-ABV99595 and ABP70049-ABP70149), where X is any number from 1 to 48. The NOVX proteins and coding sequences are useful in the manufacture of a medicament for treating a syndrome associated with a human disease, preferably a NOVX-associated disorder. The NOVX coding sequences and proteins are useful for treating, preventing or diagnosing diseases such as metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's disease, immune disorders, haematopoietic disorders, cardiovascular disorders, infertility, bronchial asthma, AIDS, dyslipidemia, metabolic disturbances associated with obesity, metabolic syndrome X or wasting disorders associated with chronic diseases or various cancers. The NOVX coding sequences and proteins may also be used as targets for the identification of small molecules that modulate or inhibit e.g. neurogenesis, cell differentiation, cell proliferation, haematopoiesis, wound healing and angiogenesis, in gene therapy, in generation of antibodies that bind immunospecifically to NOVX substances for use in therapeutic or diagnostic methods	QY	2895	EYTCEVEASKSTASLHV	EEXANCFTTEELTNLOVEEKGTA	VFTCKTEHPAA	TVTW	RKGLLE 2954
XX		Db	301	EYTCEVEASKSTASLHV	EEXANCFTTEELTNLOVEEKGTA	VFTCKTEHPAA	TVTW	RKGLLE 360

Sequence 4675 AA;

Query Match 25.5%; Score 10506; DB 5; Length 4675;

Best Local Similarity 56.1%; Pred. No. 0;

Matches 2252; Conservative 45; Mismatches 129; Indels 1586; Gaps 35;

QY 3887 ----- 3886  
Db 1441 IHGLAMADNGVYSCVCGQERTSATLTVRALPARFIEDMRNQKATEGATVTLQCKLRKAAP 1500  
QY 3887 ----- 3886  
Db 1501 VEWKGPNTLKDGRYSLKQDGTSCELQIRGLVIADAGEYSCICEQERTSATLTVRALPA 1560  
QY 3887 ----- 3886  
Db 1561 RPIEDVRNHEATEGATAVLOCELSKAAPEVWRKGSSETLRDGRYSLRQDGTGRCELQIRGL 1620  
QY 3887 ----- 3886  
Db 1621 AVEDTGEYLCVCGQERTSATLTVRALPARFIDNMTNQBAREGATATLHCELSKVAPVEWR 1680  
QY 3887 ----- 3886  
Db 1681 KGPETLRDGRHSLEQDGRSCELQIRGLAVVDAGEYSCVCGQERTSATLTVRALPARFIE 1740  
QY 3887 ----- 3886  
Db 1741 DVQNEAREGATAVLOCELSKAAPEVWRKGSSETLRDGRYSLRQDGTGRCELQIRGLSVAD 1800  
QY 3887 ----- 3886  
Db 1801 TGEYSCVCGQERTSATLTVRALPARFTQDLTKTEASEGATATLOCELSKVAPVEWKGP 1860  
QY 3887 ----- 3886  
Db 1861 TLRDGRYSLKQDGTGRCELQIRGLSVADAGEYSCMCGQERTSATLTVRDCHTLVMPHY 1920  
QY 3887 ----- 3886  
Db 1921 FQLPGLKEPBETLIYIQIPSPVILFTEGURNEAMEGATATLOCELSKAAPEVWRKGLE 1980  
QY 3887 ----- 3886  
Db 1981 ALRDGKYSLRQDGAVCLEQIRGLAMADNGVYSSLPARFIEDMRNQKATEGATVTLQCKL 2040  
QY 3887 ----- 3886  
Db 2041 RKAAPVWRKGPNTLKDGRYSLKQDGTSCELQIRGLVIADAGEYSCICEQERTSATLTV 2100  
QY 3887 ----- 3886  
Db 2101 RALPARFIEDVRNHEATEGATAVLOCELSKAAPEVWRKGSSETLRDGRYSLRQDGTGRCEL 2160  
QY 3887 ----- 3886  
Db 2161 QIRGLAVEDTGEYLCVCGQERTSATLTVRALPARFIDNMTNQBAREGATATLHCELSKVA 2220  
QY 3887 ----- 3886  
Db 2221 PVWRKGPETLRDGRHSLRQDGTGRCELQIRGLSVADAGEYSCVCGQERTSATLTVRALP 2280  
QY 3887 ----- 3886  
Db 2281 AKFTKGLRNEATEGATAMLOCELSKVAPVWRKGPETLRDGRYSLRQDGTGRCELQIRGL 2340  
QY 3887 ----- 3886  
Db 2341 LSVADTGEYSCVCGQERTSATLTVKAPQVPFRLQAEESTATLOCELSEPTATVV 2400  
QY 3922 WSKGGLQLANGREPRLOCTAELVLQDLQREDTGEYTCGQSQATSATLTVTAAPVRF 3981  
Db 2401 WSKGGLQLANGREPRLOCTAELVLQDLQREDTGEYTCGQSQATSATLTVTAAPVRF 2460  
QY 3982 LRELQHOEVDGGTAHLCCELSRAGASVWRKGSLSQLFPCAKYQMVQDGAABELLVRGVE 4041  
Db 2461 LRELQHOEVDGGTAHLCCELSRAGASVWRKGSLSQLFPCAKYQMVQDGAABELLVRGVE 2520  
QY 4042 QEDAGDYTCDTGHTQTSNASLUSVRPFRPKFKTRILQSLQEQETGDIARLCCQLSDAESGAVVQ 4101

Db 2521 QEDAGDYTCDTGHTQTSNASLUSVRGGR-----GAA-- 2549  
QY 4102 WLKEGVELHAGPKYEMRSQGATRELLIHOLEAKDTGEYACVTGGOKTAASLRVTEPEVTI 4161  
Db 2550 -----CGPOVRAAQAAGATRELLIHOLEAKDTGEYACVTGGOKTAASLRVTEPEVTI 2600  
QY 4162 VRGLVDAEVTADEDEFFSCVSRAGATGVQWCLQGLPLQSNVEVAVVRDGRHITRLKLG 4221  
Db 2601 VRGLVDAEVTADEDEFFSCVSRAGATGVQWCLQGLPLQSNVEVAVVRDGRHITRLKLG 2660  
QY 4222 VTPEDAGTVSFHLGNHASAQLTVRAPEVTILEPLQDVOLSEGQDASFOCKLSRASGOEA 4281  
Db 2661 VTPEDAGTVSFHLGNHASAQLTVRAPEVTILEPLQDVOLSEGQDASFOCKLSRASGOEA 2701  
QY 4282 RWALGCVPLQANEMNDITVEQGTLLHLLHKVTLEDACTVSVFHVGTCSSEALQKVT--- 4337  
Db 2702 -----GVPLQANEMNDITVEQGTLLHLLHKVTLEDACTVSVFHVGTCSSEALQKVT--- 2756  
QY 4338 ----- 4337  
Db 2757 CLVRGLQNVDPAGEVATFSCBDGPQSAIAVRDGIHFHSLMLSLGLGVADSGTVIFRAGPLV 2816  
QY 4338 ----- 4337  
Db 2817 STAKLLIKDPVVVVVSAMQDLAVBEGGSABELLCOYSRVPQATWMDREVTGDRHVIIE 2876  
QY 4338 -----AKNTVVRGLENVEALEGEALFE 4360  
Db 2877 QDMNVARLTFRPAIPCDSGIYSCBAAGTRVVALLQVQAKNTVVRGLENVEALEGEALFE 2936  
QY 4361 COLSPEVAANTWLLDDPEVPTSENAAEVVPPENGLRHLLILKILKLPQDSCHVTFLAGDMV 4420  
Db 2937 COLSPEVAANTWLLDDPEVPTSENAAEVVPPENGLRHLLILKILKLPQDSCHVTFLAGDMV 2996  
QY 4421 TSAFLTVR-----GWRLEILEPLKNAAVRAGAQAARFTCTLSEAVPVGEASVINGAA 4472  
Db 2997 TSAFLTVRGDCAVLVQGRLEILEPLKNAAVRAGAQAARFTCTLSEAVPVGEASVINGAA 3056  
QY 4473 VOPDDSDVTVADGSHQALLRSQPHHAGEVTFACRDVAVASARLTVLGLPDPPEDAEVV 4532  
Db 3057 VOPDDSDVTVADGSHQALLRSQPHHAGEVTFACRDVAVASARLTVLGLPDPPEDAEVV 3116  
QY 4533 AHSHTVTLNAAWPMSDGGGLCGYRVEVEKEGATQWRLCHELVGPECVVVDGLAPGETY 4592  
Db 3117 ARSHTVTLNAAWPMSDGGGLCGYRVEVEKEGATQWRLCHELVGPECVVVDGLAPGETY 3176  
QY 4593 RFRVAAGPVGAGBPFVHLPTVRLAEPKVPVPPQPSAPESRQVAAAGBDEVLSLEVVAAEAG 4652  
Db 3177 RFRVAAGPVGAGBPFVHLPTVRLAEPKVPVPPQPSAPESRQVAAAGBDEVLSLEVVAAEAG 3235  
QY 4653 EVIWHKGMERLOPGRPEVVSQGRQOMLVKFTAEQDQGEVHCGLAOGSICPAAATFOVA 4712  
Db 3236 EVIWHKGMERLOPGRPEVVSQGRQOMLVKFTAEQDQGEVHCGLAOGSICPAAATFOVA 3295  
QY 4713 LSPASVDEAPQPSLPPEAAQEGDLHLLWEALARKRMSREPTLDSISELPEEDGRSRLP 4772  
Db 3296 LSPASVDEAPQPSLPPEAAQEGDLHLLWEALARKRMSREPTLDSISELPEEDGRSRLP 3355  
QY 4773 QEAEVAPDLSEGYSTADELARTGDADLSHTSSDDESRAGTSPSLVTLKAGRGTSPLA 4832  
Db 3356 QEAEVAPDLSEGYSTADELARTGDADLSHTSSDDESRAGTSPSLVTLKAGRGTSPLA 3415  
QY 4833 SKVGAP-----AAPSVKPQOQOEPL 4852  
Db 3416 SKVSPPNLACKERFPTFRAGSLILGFGADPAPPGSERSARCTRRCAAPPPRESLKEP- 3474  
QY 4853 AAVRPPGLDLSTKJG-----DPSMD-----KAAVKIQ-----AAFK 4884  
Db 3475 ASCJF--GAEVAVELARKLOEAEATCSICLDYFTDPVMTTCGHNFRCRACIQLSWEKARGKK 3532  
QY 4885 GYKVRK-----ENKQOEGPMFSTFGDTAQVGDALRLCEVAVASADVARWLKDG 4935

Db 3533 GRKRKGSFPCEKREMSPOKLLPN-----RLTKVAEMA----- 3568  
 QY 4936 VELTDGRHHHDQLGDCGTCSLIAGLORADAGCYTCQVSNK----FCQVTHSACVVVSGS 4991  
 Db 3569 -----QQH-----FQLOKQD-----LCQEHHEPLKLFCKQDQSPICVVCRE 3604  
 QY 4992 ESEASSSGGELDDAFRAARRHLRFRKTSAPAEVSDEELFSLADEG----- 5038  
 Db 3605 SRE-----HRLHVL-----PAEEAVQGYKLEEDMEVYLREQITRTGN 3643  
 QY 5039 --PAPEPAPDWO--TYREDEFICIRFEAL-----TEARQAVTRFOEMF 5079  
 Db 3644 LQAREEQSLAEQWQKVRERRRIVLEPEKMNLYLVEEQRLLQALETEEBETASRLRESV 3703  
 QY 5080 ATL--GIGVRIKIVE-----QGPRR-----VENCISK-- 5104  
 Db 3704 ACLDRQGHSLLELLQLLEERTQGLQMLQDMKEPLSRAALLVLIHGMLNVEPPVSLP 3763  
 QY 5105 -----ETPAVVPPEPLPSLITSDAAP-----VPLTE 5131  
 Db 3764 SPLYLIAKATQLGPGTTFDPECTPLPISPP-PPPS--TEDVVPDATSAYPYLLLYE 3820  
 QY 5132 LQNEQVQDGP-----VSFDQVVTGQPMPSVRWPKDGKLLLEDDHY 5172  
 Db 3821 SRQRYLGSSPEGSGFCSDRFVAPCAV-QGTA-----FSSGR-----HY 3860

RESULT 6  
 ABP59227  
 ID ABP58227 standard; protein; 2328 AA.  
 XX  
 AC ABP58227;  
 XX  
 DT 31-MAR-2003 (first entry)  
 XX  
 DE Human cell adhesion and extracellular matrix protein 4.  
 KW Cell adhesion and extracellular matrix protein 4; CADECM-4; human;  
 KW anti-Hiv; virucide; antiallergic; antiinflammatory; antianaemic;  
 KW antiparkinsonian; nootropic; anticonvulsant; antiinfertility;  
 KW antiarteriosclerotic; antiasthmatic; immunosuppressive; antithyroid;  
 KW cytostatic; hepatotropic; dermatological; antidiabetic; nephrotropic;  
 KW antiagout; thyromimetic; neuroprotective; osteopathic; antiarthritis;  
 KW antiparasitic; antihelminthic; antiprosclerotic; uropathic; ophthalmological;  
 KW antirheumatic; haemostatic; antibacterial; protozoacide; fungicide;  
 KW gynaecological; neurexin; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200288322-A2.  
 XX  
 PD 07-NOV-2002.  
 XX  
 PF 01-MAY-2002; 2002WO-US013874.  
 XX  
 PR 02-MAY-2001; 2001US-0289290P.  
 PR 21-MAY-2001; 2001US-0292468P.  
 PR 15-JUN-2001; 2001US-0298616P.  
 PR 28-JUN-2001; 2001US-0301672P.  
 PR 04-JAN-2002; 2002US-0345008P.  
 XX  
 PA (INCY- ) INCYTE GENOMICS INC.  
 XX  
 PI Yue H, Lee EA, Duggan BM, Thangavelu K, Honchell CD, Ding L;  
 PI Hillman JL, Baughn MR, Kallio DA, Lee S, Warren BA, Xu Y, Tran UK;  
 PI Lal PG, Thornton M, Hafalia AJA, Yao MG, Nguyen DB, Gandhi AR;  
 PI Khan FA, Walhia NK, Griffin JA, Chinn AM, Elliott VS, Ramkumar J;  
 PI Arvizu CS, Forsythe IJ;  
 XX  
 DR WPI; 2003-167112/16.  
 DR N-PSDB; AB224581.  
 XX  
 PI New human cell adhesion and extracellular matrix proteins, useful for

PT diagnosing, treating or preventing autoimmune or inflammatory disorder  
 PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,  
 PT cancer or hepatitis.  
 XX  
 PS Claim 1; Page 144-150; 178pp; English.  
 XX  
 CC The present sequence is the protein sequence of Incyte polypeptide  
 CC 7326129CD1 denoted human cell adhesion and extracellular matrix protein 4  
 CC (CADECM-4). The protein is encoded by a clone isolated from a male muscle  
 CC cDNA library. Homology searches indicate it to be a titin muscle protein.  
 CC The invention provides CADECM-1 to -11 polypeptides (see ABP58224-34) and  
 CC polynucleotides (see AB224578-88), expression vectors, host cells, and  
 CC antibodies, agonists and antagonists. These are useful for diagnosing,  
 CC treating or preventing disorders associated with aberrant expression of  
 CC CADECM, particularly cell proliferative disorders (e.g. arteriosclerosis,  
 CC atherosclerosis, cirrhosis, hepatitis, paroxysmal nocturnal  
 CC haemoglobinuria, polycythaemia vera, psoriasis, primary  
 CC thrombocytopaenia or cancer), developmental disorders (e.g. renal  
 CC tubular acidosis, anaemia or mental retardation), neurological disorders  
 CC (e.g. Alzheimer's disease, Parkinson's disease or epilepsy), reproductive  
 CC disorders (e.g. infertility or a disruption in the menstrual cycle), or  
 CC autoimmune/inflammatory disorders (e.g. AIDS, allergy, asthma, autoimmune  
 CC thyroiditis, contact dermatitis, Crohn's disease, diabetes mellitus,  
 CC glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease,  
 CC Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis,  
 CC osteoarthritis, osteoporosis, pancreatitis, Reiter's syndrome, rheumatoid  
 CC arthritis, Sjogren's syndrome, uveitis), or viral, bacterial, fungal,  
 CC parasitic, protozoal or helminthic infections  
 XX  
 SQ Sequence 2328 AA;  
 Query Match 23.8%; Score 9834.5; DB 6; Length 2328;  
 Best Local Similarity 84.6%; Pred. No. 0;  
 Matches 1941; Conservative 76; Mismatches 213; Indels 65; Gaps 8;  
 QY 2096 MEVQLSHADVDSWTRDGLRFOQGPTCHLAVRGPMTLLSLRPDSDGLMVFKAEGVHT 2155  
 Db 1 MEVQLSHADVDSWTRDGLRLQGPPTCHLAVRGPMTLLSLRPDSDGLMVFKAEGVHT 60  
 QY 2156 SARLVTELPVPSRPLQDVVTEKVKVLELSPNVVRLKDGVELRAGKTWAIAA 2215  
 Db 61 SARLVTELPVPSRPLQDVVTEKVKVLELSPNVVRLKDGVELRAGKTWAIAA 120  
 QY 2216 QGACRSLLTYRCBFADQGVVVCDAHDAQSSASVKVQ----- 2251  
 Db 121 QGACRSLLTYRCBFADQGVVVCDAHDAQSSASVKVQGRNIQIVRPLEDEVMEKDGATFS 180  
 QY 2252 -----GRTYLLIYRVLAEADAGEIQFVAENAESRA 2281  
 Db 181 CEVSHDEVPGQWFEGSKLRPTDNVRIQEGRTYLLIYRVLAEADAGEIQFVAENAESRA 240  
 QY 2282 QLRVKELPVTLPRLDKIAVEKHGVLQCVSRASQAQVRFKGSQELQPGPKYELVSDG 2341  
 Db 241 QLRVKELPVTLPRLDKIAVEKHGVLQCVSRASQAQVRFKGSQELQPGPKYELVSDG 300  
 QY 2342 LYRKLIISDVHAEDEDTYTCADGDKVTSQAFFVEEQSIITVIRGLQDVVMEPAPAFCE 2401  
 Db 301 LYRKLIISDVHAEDEDTYTCADGDKVTSQAFFVEEQSIITVIRGLQDVVMEPAPAFCE 360  
 QY 2402 TSIPSVRPPKWLIGKTVLQAGGNVGLQEGTVHRLMRLRRTCTMTGPHVHTVCKSSSR 2461  
 Db 361 TSIPSVRPPKWLIGKTVLQAGGNVGLQEGTVHRLMRLRRTCTMTGPHVHTVCKSSSR 420  
 QY 2462 LVVSDIPVVLTRPLEPRTGRELQSVVLSDFRPAKAVQWYKDDTLPSPSEKFMLEGG 2521  
 Db 421 LVVSDIPVVLTRPLEPRTGRELQSVVLSDFRPAKAVQWYKDDTLPSPSEKFMLEGG 480  
 QY 2522 MAELIRLRLMPADAGVYRCQAGSAHSTVTVAREVTVTGPLQDAEAEEGHASFCE 2581  
 Db 481 MAELIRLRLMPADAGVYRCQAGSAHSTVTVAREVTVTGPLQDAEAEEGHASFCE 540  
 QY 2582 SHDEEVEWELNGMPLYNDSFHEISHKGRRTTLVLSIORADAGIVRASLKVSTGARLE 2641

Db 541 SHEDVEWNLGMPFLYNDSPHEISHKGRHRTLVLKSIQRADAGIVRASSLKVSTSRLE 600  
 Qy 2642 VRVKPVFLKALDDLSABERGTLALQCEVSDPEAHVVRKDGVOQLGPSDKYDFLHTAGTR 2701  
 Db 601 VRVKPVFLKALDDLSABERGTLALQCEVSDPEAHVVRKDGVOQLGPSDKYDFLHTAGTR 660  
 Qy 2702 GLVHDVSPEDAGLYTCHVGSEETARVRVHDLHVGITIKRLKTNVEVLGEGSCSECVLSH 2761  
 Db 661 GLVHDVSPEDAGLYTCHVGSEETARVRVHDLHVGITIKRLKTNVEVLGEGSCSECVLSH 720  
 Qy 2762 ESADPAMWTVGGKTVGSSSFQATROGRKXILVVRREAPSDAGEVVSFVRLGTSKASLI 2821  
 Db 721 ESADPAMWTVGGKTVGSSSFQATROGRKXILVVRREAPSDAGEVVSFVRLGTSKASLI 780  
 Qy 2822 VREPAALIKPLEDQWAPGDEVELRCELSDAGTFVHMLKDKAIRKSQKYDVVCEGTMA 2881  
 Db 781 VREPAALIKPLEDQWAPGDEVELRCELSDAGTFVHMLKDKAIRKSQKYDVVCEGTMA 840  
 Qy 2882 MLVIRGASLKDAGEYTCVEASKSTASLHVBEKANCETEELTNLOVEKGTAVFTCKTEH 2941  
 Db 841 MLVIRGASLKDAGEYTCVEASKSTASLHVBEKANCETEELTNLOVEKGTAVFTCKTEH 900  
 Qy 2942 PAATVVRKGLLELRASGHQPSQEGTLRLTISALEKADSDTYTCDIGQAQRAQLLVQ 3001  
 Db 901 PAATVVRKGLLELRASGHQPSQEGTLRLTISALEKADSDTYTCDIGQAQRAQLLVQ 960  
 Qy 3002 GRRVHIIEDEVDVQEGSSATFCRISPANYEPVHFWLDTPLHANELNEIDAQPGGYH 3061  
 Db 961 GRRVHIIEDEVDVQEGSSATFCRISPANYEPVHFWLDTPLHANELNEIDAQPGGYH 1020  
 Qy 3062 VLTLRQALKDQSGITYFAGQDQASALRVTEKPSVFSRELTDTATITEGEDDLTVCEST 3121  
 Db 1021 VLTLRQALKDQSGITYFAGQDQASALRVTEKPSVFSRELTDTATITEGEDDLTVCEST 1080  
 Qy 3122 CDI PMCWTKDOKTLRGSAARCOLSHEGHRQAQLLITGATLQDSGRYKCEAGACSSIVRVH 3181  
 Db 1081 CDI PMCWTKDOKTLRGSAARCOLSHEGHRQAQLLITGATLQDSGRYKCEAGACSSIVRVH 1140  
 Qy 3182 ARPVRFOALKDLEVLGGATLRCVLSSVAAPVKCYGNNVLRPGDKYSLRQEGAMLEL 3241  
 Db 1141 ARPVRFOALKDLEVLGGATLRCVLSSVAAPVKCYGNNVLRPGDKYSLRQEGAMLEL 1200  
 Qy 3242 VVRNLRPODSGRYSCSGFDQTSATLTVTALPAQFIGLKNKEATEGATATLRCESKTA 3301  
 Db 1201 VVRNLRPODSGRYSCSGFDQTSATLTVTALPAQFIGLKNKEATEGATATLRCESKTA 1260  
 Qy 3302 PVWRKGSSETLRDGRYCLRDQGMCELOIRGLAWVDAAEYSCVCGEERTSASLTIRPMP 3361  
 Db 1261 PVWRKGSSETLRDGRYCLRDQGMCELOIRGLAWVDAAEYSCVCGEERTSASLTIRPMP 1320  
 Qy 3362 AHFTGRLRHQSIEGATATLRCESKAAAPVWRKGSRLDGRHSLRQDGAVCLEQICG 3421  
 Db 1321 AHFTGRLRHQSIEGATATLRCESKAAAPVWRKGSRLDGRHSLRQDGAVCLEQICG 1380  
 Qy 3422 LAVADAGEYSCVCGEERTSASLTIVKALPAKFTGLRNEEAVEGATAMLCESKAPVW 3481  
 Db 1381 LAVADAGEYSCVCGEERTSASLTIVKALPAKFTGLRNEEAVEGATAMLCESKAPVW 1440  
 Qy 3482 RKGPENLRDGRYTLRQEGTRCELOICGLAMADAGEYLCVCGOERTSATLITRALPARFI 3541  
 Db 1441 RKGPENLRDGRYTLRQEGTRCELOICGLAMADAGEYLCVCGOERTSATLITRALPARFI 1500  
 Qy 3542 EDVKNQAREGATAVLOCELNSAAPVWRKGSSETLRDGRYSLRQDGTCKELOIRGLAMA 3601  
 Db 1501 EDVKNQAREGATAVLOCELNSAAPVWRKGSSETLRDGRYSLRQDGTCKELOIRGLAMA 1560  
 Qy 3602 DTGEYSCVCGOERTSASLTIVRALPIKFTGLRNEEATEGATAVLRCELSKQAPVWVKGH 3661  
 Db 1561 DTGEYSCVCGOERTSASLTIVRALPIKFTGLRNEEATEGATAVLRCELSKQAPVWVKGH 1620  
 Qy 3662 ETLRDGRHSLRQDGAARCELOIRGLVAEDAGEYLCMCGKERTSASLTIVRAMPSKFTIEGLR 3721  
 Db 1621 ETLRDGRHSLRQDGAARCELOIRGLVAEDAGEYLCMCGKERTSASLTIVRAMPSKFTIEGLR 1680

Qy 3722 NEEATEGDTATLWCELSKAAAPVWRKKGHETLRDGRHSLRQDGSRCCELOIRGLAVVDAGE 3781  
 Db 1681 NEEATEGDTATLWCELSKAAAPVWRKKGHETLRDGRHSLRQDGSRCCELOIRGLAVVDAGE 1740  
 Qy 3782 YSCVCGOERTSATLITVRALPARFIEDVKNQAREGATAVLOCELNSKAAAPVWRKGSSETLR 3841  
 Db 1741 YSCVCGOERTSATLITVRALPARFIEDVKNQAREGATAVLOCELNSKAAAPVWRKGSSETLR 1800  
 Qy 3842 GDRYSLRQDGRYCLRDQGMCELOIRGLHGLSVADTGEYSCVCGOERTSATLITVRAPQVPREPQSLQA 3901  
 Db 1801 GDRYSLRQDGRYCLRDQGMCELOIRGLHGLSVADTGEYSCVCGOERTSATLITVRAPQVPREPQSLQA 1860  
 Qy 3902 EBGSTATLQCELSEPTATVWSKGLQLOANGRRPRELQGGCTAEVLQDLQREDTGEYTC 3961  
 Db 1861 SEGATATLQCELSK-VAPVWRKKGPELTRDGRYSLKQDGTGRCELOIHDLVSVADAGEYSC 1919  
 Qy 3962 TCGSOATSATLITVTAAPVRFELRELQHQVDEBGTALHCELSRAGASVWRKGSLOLEPC 4021  
 Db 1920 MCGOERTSASLTIVRALPARFTEGLRNEEATEGATATLQCELSKAA-APVWRKKGLEALRDG 1978  
 Qy 4022 AKYQVQDGAALVLRGVEQEDAGDYTCDTGHTQSMASLSVVRPRPFKTRLOLEQET 4081  
 Db 1979 DKYSLRQDGAVCLELOIHGLAMADNGVYSCVCGOERTSATLITVRALPARFIEDMRNQKATE 2038  
 Qy 4082 GDIALCCOLSDAESGAVVQWIKESVELHAGPKYEMRSQGTARELLIHOLEAKDTGEYAC 4141  
 Db 2039 GATVTLQCKLRKA---APVWRKGPNTLKDGRYSLKQDGTSCLEQIRGLVIADAGEYSC 2095  
 Qy 4142 VTGGOKTAASLRVTEPVTIVRGLVDAEVTADEVEFCEVSRAGATGVQVCLQGLPLQOS 4201  
 Db 2096 ICEQERTSATLITVRALPARFIEDVNRHEATEGATAVLOCELNSKAA--VWRKGSSETLRD 2153  
 Qy 4202 NEVTEVAVRDGRIHTLRKGVTPEDAGTVSFHLGNHASSAOLTVRAPVETILEPQDVQL 4261  
 Db 2154 GDRYSLR-QDGTGRCELOIRGLAVEDEYVLCVCGOERTSATLITVRALPARFIDNNTQEA 2212  
 Qy 4262 SGGQDASFOCLRSRAGSQAARWALGGVPLQANEMNDITVEOGTLLHLLTHKVTLEDAGTV 4321  
 Db 2213 REGATATLHCELSKAAAPVE--WRKGSRLRQDGRHSLR-QDGAVCLELOICGLAVADAGEY 2269  
 Qy 4322 SFHVTCSSSEALKV 4336  
 Db 2270 SCVCGEERTSATLITV 2284  
 RESULT 7  
 ID AAO15372 standard; protein; 1665 AA.  
 XX AC AAO15372;  
 XX XX AC AAO15372;  
 DT 19-SEP-2002 (first entry)  
 XX DE Human myosin light chain kinase subfamily-related kinase protein.  
 XX KW Human; gene therapy; chromosome 1; kinase protein;  
 KW myosin light chain kinase subfamily; kinase protein-mediated disease;  
 KW transgenic animal.  
 XX OS Homo sapiens.  
 XX PN WO200240683-A2.  
 XX PD 23-MAY-2002.  
 XX PF 22-OCT-2001; 2001WO-US032616.  
 XX PR 14-NOV-2000; 2000US-00711134.  
 XX PR 17-MAY-2001; 2001US-00858664.  
 XX PA (PEKE ) PE CORP NY.  
 XX XX

Wei M, Ketchum K, Di Francesco V, Beasley EM;

WPI; 2002-500223/53.

N-PSDB; AAL43908, AAL43909.

New kinase proteins related to myosin light chain kinase subfamily and encoding polynucleotide, useful for diagnosing, treating disease or condition mediated by the kinase protein and for identifying modulators.

Claim 1: Fig 2: 96pp: English.

The invention comprises the amino acid and coding sequences (located on chromosome 1) of a human kinase protein that is related to the myosin light chain kinase subfamily. The human kinase DNA and protein sequences of the invention are useful for identifying agents that modulate the activity of the human kinase protein. Kinase-modulating agents are useful for treating a disease or condition mediated by a human kinase protein. The human kinase DNA sequences can be used to produce transgenic animals which are useful for studying the function of kinase proteins and identifying/evaluating modulators of kinase protein activity. The present amino acid sequence represents the human kinase protein of the invention

Sequence 1665 AA:

Query Match 20.4%; Score 8423; DB 5; Length 1665;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1612; Conservative 0; Mismatches 1; Indels 0

6356	QY	PPSMQVTIEDVQAGTGTAGTAFEAILI	BGDPOFSVTWYKDSVOLVDSTLS	QOQEGTYSVLV	6415
53	Db	PPSMQVTIEDVQAGTGTAGTAFEAILI	BGDPOFSVTWYKDSVOLVDSTLS	QOQEGTYSVLV	112
6416	QY	LRHVASKDAGVYTCLAQNTGGQVLCKAELLVL	GGDNEPDESEKSHRKLHSFVYKVEIG	6475	
113	Db	LRHVASKDAGVYTCLAQNTGGQVLCKAELLVL	GGDNEPDESEKSHRKLHSFVYKVEIG	172	
6476	QY	RGVFGFKRYOHNKGNKILCAAKFIPLASRTRAQAYRERDILAA	LSHPLVTGLLDQFETR	6535	
173	Db	RGVFGFKRYOHNKGNKILCAAKFIPLASRTRAQAYRERDILAA	LSHPLVTGLLDQFETR	232	
6536	QY	TLILILELCSSEBLLDRLYRGVVTAEVVKVY	QQLVEGLHYLHSHGVHLHDIKPSNLM	6595	
233	Db	TLILILELCSSEBLLDRLYRGVVTAEVVKVY	QQLVEGLHYLHSHGVHLHDIKPSNLM	292	
6596	QY	VHPAREDIKICDFGFAQNIITPAELQFSQYGSPEFVSP	PEIICQNPVSEASDIWANGVYSYL	6655	
293	Db	VHPAREDIKICDFGFAQNIITPAELQFSQYGSPEFVSP	PEIICQNPVSEASDIWANGVYSYL	352	
6656	QY	SLTCSSPFPAGESDRATILNLVLEGRVSNSSPMAAHLSEDA	KDFIKATLQAPQAPPSAAQC	6715	
353	Db	SLTCSSPFPAGESDRATILNLVLEGRVSNSSPMAAHLSEDA	KDFIKATLQAPQAPPSAAQC	412	
6716	QY	LSHPWFUKSPABEAHINTKQFKLLARSRWQSRMSYKSIL	VMRSIPELLRGPPDSPS	6775	
413	Db	LSHPWFUKSPABEAHINTKQFKLLARSRWQSRMSYKSIL	VMRSIPELLRGPPDSPS	472	
6776	QY	LGVARHLCRDTGGSSSSSSSDNELAPFAKASLPPSPVTHS	PLLHPRGFIRPSASLPEE	6835	
473	Db	LGVARHLCRDTGGSSSSSSSDNELAPFAKASLPPSPVTHS	PLLHPRGFIRPSASLPEE	532	
6836	QY	AEASERSTEAPAPPASPEGAGPPAAQCVRPHSVIRSLF	THOAGESPEHGALAPGSRHRP	6895	
533	Db	AEASERSTEAPAPPASPEGAGPPAAQCVRPHSVIRSLF	THOAGESPEHGALAPGSRHRP	592	
6896	QY	ARRHLLKGGYIAGALPGLREPLMEHRVLEBEEAAAREEQAT	LLAKAPSFETALRLPASGTH	6955	
593	Db	ARRHLLKGGYIAGALPGLREPLMEHRVLEBEEAAAREEQAT	LLAKAPSFETALRLPASGTH	652	
6956	QY	LAPGHSLSLHSDSPSTPRPSEACEAQRULPSAPSGGAP	IRDMGHGPGSKOLPSTGGHPG	7015	
653	Db	LAPGHSLSLHSDSPSTPRPSEACEAQRULPSAPSGGAP	IRDMGHGPGSKOLPSTGGHPG	712	
7016	QY	TAOPERSPSPWQCQAPFCHPKQGSAPQEGCSPPHAPVAPCP	GGSPFGSCKEAPLVSS	7075	







Db 953 EVGRAPTRSSPEPTWEDIGQVLSVQIRDLSDGDAADTISLDISEVDPAAYLNSLDYDI 1012  
Qy 7316 KYLPFERWIFRKVPKSAQPPSPMAEEELAEFPETWPGELPGHAGLEITEESDND 7375  
Db 1013 KYLPFERWIFRKVPKSAQPPSPMAEEELAEFPETWPGELPGHAGLEITEESDND 1072  
Qy 7376 ALLAEAAVGRKRWSSPSRSLSFHPGSHLPLDDEPAELGLRERVKASVEHLSRIILKGRPEG 7435  
Db 1073 ALLAEAAVGRKRWSSPSRSLSFHPGSHLPLDDEPAELGLRERVKASVEHLSRIILKGRPEG 1132  
Qy 7436 LEKEGPRKPEGLASFRLSGLKSWDRAPTFRLRELSDETIVVLGQSVTLACQVSAQPAQAAT 7495  
Db 1133 LEKEGPRKPEGLASFRLSGLKSWDRAPTFRLRELSDETIVVLGQSVTLACQVSAQPAQAAT 1192  
Qy 7496 WSKDGAPELESSRVLISATLKNFQLLILVVAEDLGYTCVSNALGTITTTGVLRKAE 7555  
Db 1193 WSKDGAPELESSRVLISATLKNFQLLILVVAEDLGYTCVSNALGTITTTGVLRKAE 1252  
Qy 7556 RPSSSPCPDICEVADGVLLVWKPVESYGPVTYIVQCSLEGGSWTTILASDIFDCCYLTSK 7615  
Db 1253 RPSSSPCPDICEVADGVLLVWKPVESYGPVTYIVQCSLEGGSWTTILASDIFDCCYLTSK 1312  
Qy 7616 LSRGGTYTFRACVSKAGMGYSYSPSQVLLGGPSHLASBEESQGRSAQPLPSTKTPAQ 7675  
Db 1313 LSRGGTYTFRACVSKAGMGYSYSPSQVLLGGPSHLASBEESQGRSAQPLPSTKTPAQ 1372  
Qy 7676 TQIORGRFSVVRQCEKASGRALAAKIPYHPKDKTAVLREYALKGLRHPHQAQHAAY 7735  
Db 1373 TQIORGRFSVVRQCEKASGRALAAKIPYHPKDKTAVLREYALKGLRHPHQAQHAAY 1432  
Qy 7736 LSPRHLVLILCLSGPELLCLAEASYSSEVKDYLMQWLSATQYLHNOHILHLDRSE 7795  
Db 1433 LSPRHLVLILCLSGPELLCLAEASYSSEVKDYLMQWLSATQYLHNOHILHLDRSE 1492  
Qy 7796 NMITEYNLLKVDLGNQASLSQEKVLPSEKFDYLETMAPELLEGOGAVPQTDIMAIGV 7855  
Db 1493 NMITEYNLLKVDLGNQASLSQEKVLPSEKFDYLETMAPELLEGOGAVPQTDIMAIGV 1552  
Qy 7856 TAFIMLSAEYFVSSEGARDLQRLKGLVRLSRVYAGLSGGAVAFRLSTLCAQPWGRPCA 7915  
Db 1553 TAFIMLSAEYFVSSEGARDLQRLKGLVRLSRVYAGLSGGAVAFRLSTLCAQPWGRPCA 1612  
Qy 7916 SSCLOCPWLTEGAPACSPAPVPTPTARLRFVVRNREKRALLYKRNHLAQVR 7968  
Db 1613 SSCLOCPWLTEGAPACSPAPVPTPTARLRFVVRNREKRALLYKRNHLAQVR 1665

RESULT 9  
AAB85504  
ID AAB85504 standard; protein; 1618 AA.  
XX AC AAB85504;  
XX DT 25-SEP-2001 (first entry)  
XX DE Human protein kinase SGK145.

Protein kinase; enzyme; cytosolic; neurotropic; neuroprotective; human;  
antiparkinsonian; virucide; antibacterial; antifungal; antimigraine;  
analgesic; hypotensive; hypertensive; immunosuppressive; antiallergic;  
antipsoriatic; antirheumatic; antiarthritic; ophthalmological; anorectic;  
osteopathic; thrombolytic; antiarteriosclerotic; antiasthmatic;  
vasotropic; antidiabetic; gene therapy.

OS Homo sapiens.  
XX WC200155356-A2.  
XX PD 02-AUG-2001.  
XX PF 25-JAN-2001; 2001WO-US002337.

PR 25-JAN-2000; 2000US-0178079P.  
PR 31-JAN-2000; 2000US-0179364P.  
PR 17-FEB-2000; 2000US-0183173P.  
PR 17-MAR-2000; 2000US-0190162P.  
PR 29-MAR-2000; 2000US-0193404P.  
PR 13-NOV-2000; 2000US-0247013P.  
XX (SUGF-) SUGEN INC.  
PA Plowman G, Whyte D, Manning G, Sudarsanam S, Martinez R;  
PI WPI; 2001-476202/51.  
XX N-PSDB; AAH46904.  
DR Kinase polypeptides useful for treating cancers, Alzheimer's disease,  
XX viral infections, diabetes, obesity, organ transplant rejection and  
XX rheumatoid arthritis.  
PS Claim 7; Page 215; 218pp; English.  
XX The invention provides human protein kinases and protein kinase-like  
XX enzymes and polynucleotides encoding the polypeptides. The kinase  
CC polypeptides and their modulators are useful for treating a disease or  
CC disorder such as cancer, immune-related diseases, cardiovascular disease,  
CC brain or neuronal-associated disease and metabolic disorders, including  
CC cancers of tissues, cancers of hematopoietic origin, diseases of the  
CC central nervous system, diseases of the peripheral nervous system,  
CC Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic  
CC lateral sclerosis, viral infections, infections caused by prions,  
CC bacteria and fungi, ocular diseases, migraines, pain, sexual dysfunction,  
CC mood disorders, attention disorders, neurological disorders, hypotension,  
CC hypertension, psychotic disorders, organ transplant rejection. They are also useful  
CC for treating rhinitis, autoimmunity, atherosclerosis, psoriasis,  
CC osteoarthritis, asthma, chronic inflammatory pelvic disease, chronic  
CC inflammatory bowel disease, rheumatoid arthritis, metabolic disorders  
CC such as diabetes, obesity, cardiovascular diseases such as reperfusion  
CC injury, coronary thrombosis, clotting disorders and atherosclerosis.  
CC ocular diseases such as glaucoma, retinopathy and macular degeneration,  
CC psychiatric and neurological disorders such as anxiety, schizophrenia,  
CC dementia, manic depression, etc. The polynucleotides are useful in gene  
CC therapy techniques to treat the above mentioned disorders. Sequences  
CC AAB85491-85522 represent the human protein kinases of the invention  
XX Sequence 1618 AA;  
SQ

Query Match 20.4%; Score 8407; DB 4; Length 1618;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1612; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

Qy 6357 PSMQVTIEDVQAQTGTAQFEALIEGDGPPSVTWYKDSVQVLDSTLSQCGTTVSLVL 6416  
Db 1 PSMQVTIEDVQAQTGTAQFEALIEGDGPPSVTWYKDSVQVLDSTLSQCGTTVSLVL 60  
Qy 6417 RHVASKDAGVYTCIAQNTGGQVLCKAELLVLG-----GDNPPDSEKQSHRKLHSFYEV 6470  
Db 61 RHVASKDAGVYTCIAQNTGGQVLCKAELLVLGAAASHSLGDNPPDSEKQSHRKLHSFYEV 120  
Qy 6471 KEBIGRGVGFVVRVQHKGNKILCAAKFPLSRTRFAQAYRERDILAAISHPLVTLGLDQ 6530  
Db 121 KEBIGRGVGFVVRVQHKGNKILCAAKFPLSRTRFAQAYRERDILAAISHPLVTLGLDQ 180  
Qy 6531 FETRKTLILILELCSSEELLDRLYRGVVTAEVKVYIQQLVEGLHYLHSHGVHLIDIKP 6590  
Db 181 FETRKTLILILELCSSEELLDRLYRGVVTAEVKVYIQQLVEGLHYLHSHGVHLIDIKP 240  
Qy 6591 SNILMVHPAREDIKICDFGFAQNITPAELQFQSGYGFVSPEIIOQNVSASDIIWAMG 6650  
Db 241 SNILMVHPAREDIKICDFGFAQNITPAELQFQSGYGFVSPEIIOQNVSASDIIWAMG 300  
Qy 6651 VISYLSITCSSPAGESDRATLNLVLEGRVSVSSPMAAHLSEDAKDFIKATLQRAQARP 6710  
Db 301 VISYLSITCSSPAGESDRATLNLVLEGRVSVSSPMAAHLSEDAKDFIKATLQRAQARP 360

QY	6711	SAAQCLSHPWFLKSWPABEAHFINTKQKFLIARSRWORSIMWSYKSLVWMSIPBELLRGP	6770
DB	361	SAAQCLSHPWFLKSWPABEAHFINTKQKFLIARSRWORSIMWSYKSLVWMSIPBELLRGP	420
QY	6771	PDSPSLGVARHLCRDTGSSSSSSSSDNEIAPFAFAKSLPSPVTHSPLLHPRGFLRPSA	6830
DB	421	PDSPSLGVARHLCRDTGSSSSSSSSDNEIAPFAFAKSLPSPVTHSPLLHPRGFLRPSA	480
QY	6831	SLPBEAEASERSTBAPAPASPEGAGPPAAQCGCVPRHSVIRSLFHYOAGSPSEHGALAPG	6890
DB	481	SLPBEAEASERSTBAPAPASPEGAGPPAAQCGCVPRHSVIRSLFHYOAGSPSEHGALAPG	540
QY	6891	SRRHPARRHLLKGYTAGALPGLREPLMEHRVLEBEAAREQATLLAKAPSFETALRLP	6950
DB	541	SRRHPARRHLLKGYTAGALPGLREPLMEHRVLEBEAAREQATLLAKAPSFETALRLP	600
QY	6951	ASGTHLAPGHSHSLHSDSPSTRPSSBACGEAQLPSPAGCGAPIRDMGHIPQSGKQLPST	7010
DB	601	ASGTHLAPGHSHSLHSDSPSTRPSSBACGEAQLPSPAGCGAPIRDMGHIPQSGKQLPST	660
QY	7011	GGHPTAQPERSPDSPWQGPAPCHPKQGSAPQEGCSHPHAPACPPGSPFPGSCKEAP	7070
DB	661	GGHPTAQPERSPDSPWQGPAPCHPKQGSAPQEGCSHPHAPACPPGSPFPGSCKEAP	720
QY	7071	LVPSSPFLQGPAPAPAKASPPLDXKMGPGDISLPGRPKFGPCSPGSAQSSQVSS	7130
DB	721	LVPSSPFLQGPAPAPAKASPPLDXKMGPGDISLPGRPKFGPCSPGSAQSSQVSS	780
QY	7131	LRVGSSQVGTPEPGLDAEGWTQAEADLSSTPTLQRPQEQATMRKESLGGRCGYAGVAG	7190
DB	781	LRVGSSQVGTPEPGLDAEGWTQAEADLSSTPTLQRPQEQATMRKESLGGRCGYAGVAG	840
QY	7191	YGTFAFGDAGCMLQGCPMMARIAMAYSQSEEEQEERAEASQSEEQEERAEASPLPQVS	7250
DB	841	YGTFAFGDAGCMLQGCPMMARIAMAYSQSEEEQEERAEASQSEEQEERAEASPLPQVS	900
QY	7251	ARPYPEVGRAPTRSPSEPTPWEDIGQVSLQVIRLSDGDAEAADTISLDISEVDPAYNLIS	7310
DB	901	ARPYPEVGRAPTRSPSEPTPWEDIGQVSLQVIRLSDGDAEAADTISLDISEVDPAYNLIS	960
QY	7311	DLYDIKYLPEFEMI FRKVPKSAQPEPPSMAEEBIAEFPEPTWPMGELGPHAGLEITEE	7370
DB	961	DLYDIKYLPEFEMI FRKVPKSAQPEPPSMAEEBIAEFPEPTWPMGELGPHAGLEITEE	1020
QY	7371	SEDYDALLAEAAVCRKWKWSSPSRSLPHFPGRHPLDEPAELGLRVRKVASVHHISRLK	7430
DB	1021	SEDYDALLAEAAVCRKWKWSSPSRSLPHFPGRHPLDEPAELGLRVRKVASVHHISRLK	1080
QY	7431	GRPEGLEKEGPPRKKPGGLASFRLSGLKSWDRAPTFRLSDETVVLGQSVTLACQVSAQP	7490
DB	1081	GRPEGLEKEGPPRKKPGGLASFRLSGLKSWDRAPTFRLSDETVVLGQSVTLACQVSAQP	1140
QY	7491	AAQATWKGDAPLBSSSRVLISATLKNFOLLTILVVVAEDLGVTYCSVSNALGTVTTTGV	7550
DB	1141	AAQATWKGDAPLBSSSRVLISATLKNFOLLTILVVVAEDLGVTYCSVSNALGTVTTTGV	1200
QY	7551	LRKAERSSSPCPDIGEYADGVLLVWKPYESYGPVTYIVQCSEGGSWTTLASDFDCC	7610
DB	1201	LRKAERSSSPCPDIGEYADGVLLVWKPYESYGPVTYIVQCSEGGSWTTLASDFDCC	1260
QY	7611	YLTSLKSRGGTYTRPTACVSKAGMPYSSPSQVLLGGPSHLASBESQGRSAQPLPSTK	7670
DB	1261	YLTSLKSRGGTYTRPTACVSKAGMPYSSPSQVLLGGPSHLASBESQGRSAQPLPSTK	1320
QY	7671	TFAFQTOIQGRFSVVRQCKEASGRALAAKIIPYHPKDKTAVLRYEALKLGRHHPHLAQ	7730
DB	1321	TFAFQTOIQGRFSVVRQCKEASGRALAAKIIPYHPKDKTAVLRYEALKLGRHHPHLAQ	1380
QY	7731	LHAAYLSPRHVLVILELCSGPELLPCIAERASYSSEVKDYLWMLSATQYLHNOHILHL	7790
DB	1381	LHAAYLSPRHVLVILELCSGPELLPCIAERASYSSEVKDYLWMLSATQYLHNOHILHL	1440

Qy	7791	DLSENMIITEYNLLKVVLDIGNAQSLSQEKLPSDFKFOYLETMAPELLEGGQAVPQTDI	7850
Db	1441	DLSENMIITEYNLLKVVLDIGNAQSLSQEKLPSDFKFOYLETMAPELLEGGQAVPQTDI	1500
Qy	7851	WAIQVTAFLMSAEYPVSSGARDLQRLKGLVRLSRCVAGLSGGAVAFRLSTLCAQPW	7910
Db	1501	WAIQVTAFLMSAEYPVSSGARDLQRLKGLVRLSRCVAGLSGGAVAFRLSTLCAQPW	1560
Qy	7911	GRPCASSCLOCPWLTEGPACSPAPVTPPTARLRFVFNREKRRALLYRHNLQAVR	7968
Db	1561	GRPCASSCLOCPWLTEGPACSPAPVTPPTARLRFVFNREKRRALLYRHNLQAVR	1618
RESULT 10			
AA	B30568	standard; protein; 1610 AA.	
XX	AC		
XX	AA	B30568;	
XX	AC		
DT	19-MAR-2001	(first entry)	
XX	XX		
DE	XX	A full length human signal transduction polypeptide.	
XX	XX		
KW	XX	Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;	
KW	XX	congestive heart failure; dilated congestive cardiomyopathy;	
KW	XX	hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;	
KW	XX	mitral valve disease; aortic valve disease; tricuspid valve disease;	
KW	XX	myocardial infarction; cardiac arrhythmia; arteriosclerosis;	
KW	XX	atherosclerosis; cardiac tumour; microbial infection.	
XX	XX		
OS	XX	Homo sapiens.	
XX	XX		
PN	XX	W0200063381-A1.	
XX	XX		
PD	XX	26-OCT-2000.	
XX	XX		
PF	XX	11-APR-2000; 2000WO-US009488.	
XX	XX		
PR	XX	16-APR-1999; 99US-0129553P.	
XX	XX	(SCIO-) SCIOS INC.	
PA	XX		
PI	XX	Zeng W, Stanton L, Kong H;	
XX	XX		
DR	XX	WPI; 2001-007013/01.	
DR	XX	N-PSDB; AAC62286.	
XX	XX		
PT	XX	Novel h19G5 polypeptides capable of regulating signal transduction and	
PT	XX	exhibiting kinase activity useful for identifying antibodies to treat	
PT	XX	cardiac diseases, and additional mediators of signal transduction.	
XX	XX		
PS	XX	Claim 1; Page 61-65; 81pp; English.	
XX	XX		
CC	XX	The present sequence represents a human protein with putative function in	
CC	XX	signal transduction. The polypeptide is designated H19G5. The protein is	
CC	XX	capable of regulating signal transduction and exhibits kinase activity.	
CC	XX	The H19G5 transcript is expressed in the heart. H19G5 polypeptides and	
CC	XX	polynucleotides are useful for preventing or treating a cardiac disease,	
CC	XX	such as congestive heart failure, dilated congestive cardiomyopathy,	
CC	XX	hypertrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve	
CC	XX	disease, aortic valve disease or tricuspid valve disease, angina	
CC	XX	pectoris, myocardial infarction, cardiac arrhythmia, pulmonary, arterial	
CC	XX	or renovascular hypertension, arteriosclerosis, atherosclerosis and	
CC	XX	cardiac tumours in humans. The polypeptide is also useful for detecting	
CC	XX	the expression of a protein capable of regulating signal transduction or	
CC	XX	the expression of a protein capable of acting as a donor or acceptor	
CC	XX	molecule of a phosphate group. The monoclonal antibodies can be used as	
CC	XX	probes for detecting discrete antigens expressed by tissue or cell	
CC	XX	samples, and therefore used in humans for localization and monitoring of	
CC	XX	microbial infection	
XX	XX		
SQ	XX	Sequence 1610 AA;	

Query Match		20.4%;	Score 8403;	DB 4;	Length 1610;	
Best Local Similarity		99.98;	Pred. No. 0;			
Matches 1609;		Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;	
Qy	6359	MQVTIEDVQAGTGTAGTAF	FEAII	EGDPDQPSVTWYKDSVQLVDSTRLSQQQEGTYSVLVRH	6418	
Db	1	MQVTIEDVQAGTGTAGTAF	FEAII	EGDPDQPSVTWYKDSVQLVDSTRLSQQQEGTYSVLVRH	60	
Qy	6419	VASXDAGVYTC	LAQNTGGVLC	KAELLVLGDDNEPDSEKSHRKHLSHFYEVKEEIGRGV	6478	
Db	61	VASXDAGVYTC	LAQNTGGVLC	KAELLVLGDDNEPDSEKSHRKHLSHFYEVKEEIGRGV	120	
Qy	6479	FGFVYRQVHKGNTKILCAAKT	PIRSTR	QAYRERDILAALSHPLVTGLDDQFETRTKLI	6538	
Db	121	FGFVYRQVHKGNTKILCAAKT	PIRSTR	QAYRERDILAALSHPLVTGLDDQFETRTKLI	180	
Qy	6539	LILELCSSEELL	DLRYLKGVT	TEAEVYVYICQLVEGLHYLHSHCVLHLDIKPSNIIWVHP	6598	
Db	181	LILELCSSEELL	DLRYLKGVT	TEAEVYVYICQLVEGLHYLHSHCVLHLDIKPSNIIWVHP	240	
Qy	6599	AREDIKICDFGFAQNI	TPAEILQF	QSGSPFVSPETIIQQNPVSEASDIWANGVISYLSLT	6658	
Db	241	AREDIKICDFGFAQNI	TPAEILQF	QSGSPFVSPETIIQQNPVSEASDIWANGVISYLSLT	300	
Qy	6659	CSSPAGESDRATLLNV	LEGVSVS	SPMAHLSEDAKDPIKATLORAPQAPRAAQCLSH	6718	
Db	301	CSSPAGESDRATLLNV	LEGVSVS	SPMAHLSEDAKDPIKATLORAPQAPRAAQCLSH	360	
Qy	6719	PWFLLKSNPABEAHINT	KQLKFL	LARGWORSLSYKSIILVMSIPELLRGPDSPLGV	6778	
Db	361	PWFLLKSNPABEAHINT	KQLKFL	LARGWORSLSYKSIILVMSIPELLRGPDSPLGV	420	
Qy	6779	ARHLCRTDGTSSSSSS	SDNELAP	FAKASLPSPVTHSLPHRGPFRPSASLPBEAEA	6838	
Db	421	ARHLCRTDGTSSSSSS	SDNELAP	FAKASLPSPVTHSLPHRGPFRPSASLPBEAEA	480	
Qy	6839	SERSTEAPAPASPEGAP	PAACQV	PRHSVIRSLFYHQGESPEHGALAPGRRHPARR	6898	
Db	481	SERSTEAPAPASPEGAP	PAACQV	PRHSVIRSLFYHQGESPEHGALAPGRRHPARR	540	
Qy	6899	RHLKGGYIAGALPGL	REPLME	HRVLBEEAAREEQATLLAKAPSFETALRLPASGTHLAP	6958	
Db	541	RHLKGGYIAGALPGL	REPLME	HRVLBEEAAREEQATLLAKAPSFETALRLPASGTHLAP	600	
Qy	6959	GHSLSLEHDSST	TRPSEACGEA	QRLPSAPSGAPIRDMCHPQGSKQLSTGCHPCTAQ	7018	
Db	601	GHSLSLEHDSST	TRPSEACGEA	QRLPSAPSGAPIRDMCHPQGSKQLSTGCHPCTAQ	660	
Qy	7019	PERSPSPWQAPAPF	CHPKQGSAP	QGCSPHAPVAPCPGSPFPFGCKEAPLVPSSPFL	7078	
Db	661	PERSPSPWQAPAPF	CHPKQGSAP	QGCSPHAPVAPCPGSPFPFGCKEAPLVPSSPFL	720	
Qy	7079	GQOAPAPAKASPL	DOSKMGPGDISL	PGRPKPGCSPPGASQASSQSSVSLRVGSSQV	7138	
Db	721	GQOAPAPAKASPL	DOSKMGPGDISL	PGRPKPGCSPPGASQASSQSSVSLRVGSSQV	780	
Qy	7139	GTEPGPSLDAGWTQ	EAEDLS	DTPTLQRPQEQATMEKFSILGGSGVAGVAGTTFAGG	7198	
Db	781	GTEPGPSLDAGWTQ	EAEDLS	DTPTLQRPQEQATMEKFSILGGSGVAGVAGTTFAGG	840	
Qy	7199	DAGMGLGQGPWARI	AMAVS	QSEEEQEPARAEQSEEQEPARAEPLQVARSAPPEVG	7258	
Db	841	DAGMGLGQGPWARI	AMAVS	QSEEEQEPARAEQSEEQEPARAEPLQVARSAPPEVG	900	
Qy	7259	RAPTRSSPEPTWED	IGOVSLVQIR	DLSDGAERADTISLDSIVDPAYLNLSDLYDKYL	7318	
Db	901	RAPTRSSPEPTWED	IGOVSLVQIR	DLSDGAERADTISLDSIVDPAYLNLSDLYDKYL	960	
Qy	7319	PFEFMIFRKVPKSAQ	PEPPSPWABEE	LAEPPEPTWMPGELGPHAGLEITEESDVEDDALL	7378	
Db	961	PFEFMIFRKVPKSAQ	PEPPSPWABEE	LAEPPEPTWMPGELGPHAGLEITEESDVEDDALL	1020	
Qy	7379	AEAAVGRKRRKWS	SPSRSLFHP	PGRHLPLDEPAELGLRERVKASVEHISRIILKGRPEGLEK	7438	

Db	1021	AEAAVGRKRRKWS	SPSRSLFHP	PGRHLPLDEPAELGLRERVKASVEHISRIILKGRPEGLEK	1080	
Qy	7439	EGPPRRKKGCLAS	FRLSGLKSWDRAPT	FLRELSDETVVLGQSVTLACQVSAQPAQAATWSK	7498	
Db	1081	EGPPRRKKGCLAS	FRLSGLKSWDRAPT	FLRELSDETVVLGQSVTLACQVSAQPAQAATWSK	1140	
Qy	7499	DCAPLESSSRVLIS	ATLKNFQ	LLTVLVVAEDLGYVTCVSNALGTVTTTGVLRKAERPS	7558	
Db	1141	DCAPLESSSRVLIS	ATLKNFQ	LLTVLVVAEDLGYVTCVSNALGTVTTTGVLRKAERPS	1200	
Qy	7559	SSPCPDIGEVADG	VLLWKP	VESGPTYYIVQCSLEGSWTTLASDIFDCCLTSKLSR	7618	
Db	1201	SSPCPDIGEVADG	VLLWKP	VESGPTYYIVQCSLEGSWTTLASDIFDCCLTSKLSR	1260	
Qy	7619	GGTYIFRTACTV	SKAGMG	PISSPSEQVLLGAPSHLASEESQGRSAQPLPSKTFQI	7678	
Db	1261	GGTYIFRTACTV	SKAGMG	PISSPSEQVLLGAPSHLASEESQGRSAQPLPSKTFQI	1320	
Qy	7679	QGRGSRVVRQ	CKEASGRALA	AKIIPYHPKOKTAVLREYEAALGLRPHLAQLHAAYLSP	7738	
Db	1321	QGRGSRVVRQ	CKEASGRALA	AKIIPYHPKOKTAVLREYEAALGLRPHLAQLHAAYLSP	1380	
Qy	7739	RHLVILELCSP	PELLPCLABE	ASYSSEVKDYLWQMLSATQYLHNOHILHLDLSENMI	7798	
Db	1381	RHLVILELCSP	PELLPCLABE	ASYSSEVKDYLWQMLSATQYLHNOHILHLDLSENMI	1440	
Qy	7799	ITEYNLLKVVD	LGNQ	ASLSQEKVLPDKFYLETMAPELLEGQGA	7858	
Db	1441	ITEYNLLKVVD	LGNQ	ASLSQEKVLPDKFYLETMAPELLEGQGA	1500	
Qy	7859	IMLSAEYPVSS	SGARDLQ	RLKGLVRLSRCVAGLSGGAVAFRLSTLCAQPKGRPCASSC	7918	
Db	1501	IMLSAEYPVSS	SGARDLQ	RLKGLVRLSRCVAGLSGGAVAFRLSTLCAQPKGRPCASSC	1560	
Qy	7919	LQCPWLTEBGP	ACSRPAPVT	FPTARLVFVRNRRKRRALLYKRNHLAQVR	7968	
Db	1561	LQCPWLTEBGP	ACSRPAPVT	FPTARLVFVRNRRKRRALLYKRNHLAQVR	1610	

RESULT 11

AAB30567  
ID AAB30567 standard; protein; 1351 AA.  
XX  
AC AAB30567;  
XX  
DT 19-MAR-2001 (first entry)  
XX  
DE Amino acid sequence of a human signal transduction polypeptide.  
XX  
KW Signal transduction; H19C5; kinase; cardiac disease; angina pectoris;  
KW congestive heart failure; dilated congestive cardiomyopathy;  
KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;  
KW mitral valve disease; aortic valve disease; tricuspid valve disease;  
KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;  
KW atherosclerosis; cardiac tumour; microbial infection.  
XX  
OS Homo sapiens.  
XX  
PN WO200063381-A1.  
XX  
PD 26-OCT-2000.  
XX  
PF 11-APR-2000; 2000WO-US009488.  
XX  
PR 16-APR-1999; 99US-0129553P.  
XX  
PA (SCIO-) SCIOS INC.  
XX  
PI Zeng W, Stanton L, Kong H;  
XX  
DR WPI: 2001-007013/01.  
DR N-PSDB; AAC62285.

XX Novel h19G5 polypeptides capable of regulating signal transduction and  
PT exhibiting kinase activity useful for identifying antibodies to treat  
PT cardiac diseases, and additional mediators of signal transduction.  
XX  
PS Claim 1; Page 55-57; 81pp; English.  
XX  
CC The present sequence represents a human protein with putative function in  
CC signal transduction. The polypeptide is designated H19G5. The protein is  
CC capable of regulating signal transduction and exhibits kinase activity.  
CC The H19G5 transcript is expressed in the heart. H19G5 polypeptides and  
CC polynucleotides are useful for preventing or treating a cardiac disease,  
CC such as congestive heart failure, dilated congestive cardiomyopathy,  
CC hypertrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve  
CC disease, aortic valve disease or tricuspid valve disease, angina  
CC pectoris, myocardial infarction, cardiac arrhythmia, pulmonary, arterial  
CC or renovascular hypertension, arteriosclerosis, atherosclerosis and  
CC cardiac tumours in humans. The polypeptide is also useful for detecting  
CC the expression of a protein capable of regulating signal transduction or  
CC the expression of a protein capable of acting as a donor or acceptor  
CC molecule of a phosphate group. The monoclonal antibodies can be used as  
CC probes for detecting discrete antigens expressed by tissue or cell  
CC samples, and therefore used in humans for localization and monitoring of  
CC microbial infection  
XX  
SQ Sequence 1351 AA;

Query Match 17.1%; Score 7073; DB 4; Length 1351;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1349; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6619 LQFQYQSPVSPVPEIIQQNPVSPASDIWANGVTSYLSLTCSPFAGESDRAILLNVLEG 6678  
DB 2 VQFQYQSPVSPVPEIIQQNPVSPASDIWANGVTSYLSLTCSPFAGESDRAILLNVLEG 61

QY 6679 RVWSNPMAAHLSDEKDFIKATLQAPQAPRPAQAQCLSHPWFLKMPABEAHFINTKQL 6738  
DB 62 RVWSNPMAAHLSDEKDFIKATLQAPQAPRPAQAQCLSHPWFLKMPABEAHFINTKQL 121

QY 6739 KFLARSRWORSLSYKSIIVMSRIPPELLRGPDSPSLGVARHLCDTGGSSSSSSSSN 6798  
DB 122 KFLARSRWORSLSYKSIIVMSRIPPELLRGPDSPSLGVARHLCDTGGSSSSSSSSN 181

QY 6799 ELAPFARAKSLPPSPVTHPLHPRGFLRPSASLPEEAESERSTAPAPPASPEGAGPP 6858  
DB 182 ELAPFARAKSLPPSPVTHPLHPRGFLRPSASLPEEAESERSTAPAPPASPEGAGPP 241

QY 6859 AAQCVPBHSVIRSLFYHQGESPEHGALAPGRRHPARRRHLLKGYTAGALPGLREPL 6918  
DB 242 AAQCVPBHSVIRSLFYHQGESPEHGALAPGRRHPARRRHLLKGYTAGALPGLREPL 301

QY 6919 MEHVLSEEAAREQAALLAKAPSFETALRLPASGTHLAPGHSLSLEHDSPTSPRPSSEA 6978  
DB 302 MEHVLSEEAAREQAALLAKAPSFETALRLPASGTHLAPGHSLSLEHDSPTSPRPSSEA 361

QY 6979 CGEQRILPSPAGAPTRDMCHQPGSKQLPSTGCHPGTAQPERPSPDPSWGQAPPCHPK 7038  
DB 362 CGEQRILPSPAGAPTRDMCHQPGSKQLPSTGCHPGTAQPERPSPDPSWGQAPPCHPK 421

QY 7039 QGSAPOGCGPHPAVAPCPGSPFPQCKEAPLVPSPFLGQAPPAKAPASPLDLSKM 7098  
DB 422 QGSAPOGCGPHPAVAPCPGSPFPQCKEAPLVPSPFLGQAPPAKAPASPLDLSKM 481

QY 7099 GPGDISLPGKXPGPCSSPGSASQASSQVSSIRVSSQVGTGTPGSLDAEGMTQBAEDL 7158  
DB 482 GPGDISLPGKXPGPCSSPGSASQASSQVSSIRVSSQVGTGTPGSLDAEGMTQBAEDL 541

QY 7159 SDSTPTLQRPQEQATMKFKSLGGRGGYAGVAGTFAFGDAGGMLGQGMWARIWAYS 7218  
DB 542 SDSTPTLQRPQEQATMKFKSLGGRGGYAGVAGTFAFGDAGGMLGQGMWARIWAYS 601

QY 7219 QSEEEQEAEARASQSEEQAEARASPLPQVSARPVPEVGRAPTRSPPTPWEDIGQVS 7278

DB 602 QSEEEQEAEARASQSEEQAEARASPLPQVSARPVPEVGRAPTRSPPTPWEDIGQVS 661  
QY 7279 LVQIRDLSDGDAEADTISLDISEVDPAVNLSDLYDIKYLPEFEMIFRKPFSQAQPPPS 7338  
DB 662 LVQIRDLSDGDAEADTISLDISEVDPAVNLSDLYDIKYLPEFEMIFRKPFSQAQPPPS 721  
QY 7339 PMAEBELAEFPEPTWMPGELGPHAGUEITESESDVDALLAEAAVGRKRKWSSPRSRSLFH 7398  
DB 722 PMAEBELAEFPEPTWMPGELGPHAGUEITESESDVDALLAEAAVGRKRKWSSPRSRSLFH 781  
QY 7399 PFRHLPLDEPAELGLRERVVASVEHISRIUKGPEGLKEGPRKPKGLASFLSLGSLK 7458  
DB 782 PFRHLPLDEPAELGLRERVVASVEHISRIUKGPEGLKEGPRKPKGLASFLSLGSLK 841  
QY 7459 WDRAPTFLRELSDETIVLGQSVTLACQVSAQAQAATWSKDGAPLESRRVLIISATLKNF 7518  
DB 842 WDRAPTFLRELSDETIVLGQSVTLACQVSAQAQAATWSKDGAPLESRRVLIISATLKNF 901  
QY 7519 QLTILVAVASDLGVYTCVSNALGTVTTTGVLSKAEPRSSPCPDIGEVADGVLVWK 7578  
DB 902 QLTILVAVASDLGVYTCVSNALGTVTTTGVLSKAEPRSSPCPDIGEVADGVLVWK 961  
QY 7579 PVESYGPVTVYVQCSLEGSSWTTLASDIFDCCYLTSLSRGTYTFTTACVSKAGMGFYS 7638  
DB 962 PVESYGPVTVYVQCSLEGSSWTTLASDIFDCCYLTSLSRGTYTFTTACVSKAGMGFYS 1021

QY 7639 SPSEQVILGGPSHLASBEESQGRSAQPLPSTKTFATQTOIGRFSVVRQWEXASGRAL 7698  
DB 1022 SPSEQVILGGPSHLASBEESQGRSAQPLPSTKTFATQTOIGRFSVVRQWEXASGRAL 1081

QY 7699 AAKIIPYKDKTAVLREYEALKGLRHPHLAQLHAAVLSPRHLVLIILELCSGPPELLPCLA 7758  
DB 1082 AAKIIPYKDKTAVLREYEALKGLRHPHLAQLHAAVLSPRHLVLIILELCSGPPELLPCLA 1141

QY 7759 ERASYSESEVKDYLMQMLSATQYLHNOHILHLDLRSENMIITEYNLLKVDLGNASLSQ 7818  
DB 1142 ERASYSESEVKDYLMQMLSATQYLHNOHILHLDLRSENMIITEYNLLKVDLGNASLSQ 1201

QY 7819 EKVLPDSKDFKYLETMAPELLEGGQAVPQTDIWAIGVTAFTMLSAEYVPSSEGARDLQRG 7878  
DB 1202 EKVLPDSKDFKYLETMAPELLEGGQAVPQTDIWAIGVTAFTMLSAEYVPSSEGARDLQRG 1261

QY 7879 LRKGLVRLSRCYAGLSGGAVFLSTLCAQWGPBPCASSCLQCPLWTEBGPACSRPAPVT 7938  
DB 1262 LRKGLVRLSRCYAGLSGGAVFLSTLCAQWGPBPCASSCLQCPLWTEBGPACSRPAPVT 1321

QY 7939 PPTARLVFVNRKRRALLYKRNLAQVR 7968  
DB 1322 PPTARLVFVNRKRRALLYKRNLAQVR 1351

## RESULT 12

AAB30570

ID AAB30570 standard; protein; 871 AA.

XX AAB30570;

AC AAB30570;

XX 19-MAR-2001 (first entry)

XX A splice variant of a signal transduction polypeptide.

DE Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;  
KW congestive heart failure; dilated congestive cardiomyopathy;  
KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;  
KW mitral valve disease; aortic valve disease; tricuspid valve disease;  
KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;  
KW atherosclerosis; cardiac tumour; microbial infection; splice variant.

OS Homo sapiens.  
XX WO2000063381-A1.

XX 26-OCT-2000.  
PD

XX 11-APR-2000; 2000WO-US009488.  
PF 16-APR-1999; 99US-0129553P.  
XX (SCIO-) SCIOS INC.  
XX Zeng W, Stanton L, Kong H;  
PI WPI; 2001-007013/01.  
XX Novel h19G5 polypeptides capable of regulating signal transduction and  
PT exhibiting kinase activity useful for identifying antibodies to treat  
PT cardiac diseases, and additional mediators of signal transduction.  
XX  
PS Claim 1; Page 74-76; 81pp; English.  
CC The present sequence represents a splice variant of human in signal  
CC transduction polypeptide. The polypeptide is designated H19G5. The  
CC protein is capable of regulating signal transduction and exhibits kinase  
CC activity. The H19G5 transcript is expressed in the heart. H19G5  
CC polypeptides and polynucleotides are useful for preventing or treating a  
CC cardiac disease, such as congestive heart failure, dilated congestive  
CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,  
CC mitral valve disease, aortic valve disease or tricuspid valve disease,  
CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,  
CC arterial or thenovascular hypertension, arteriosclerosis, atherosclerosis  
CC and cardiac tumours in humans. The polypeptide is also useful for  
CC detecting the expression of a protein capable of regulating signal  
CC transduction or the expression of a protein capable of acting as a donor  
CC or acceptor molecule of a phosphate group. The monoclonal antibodies can  
CC be used as probes for detecting discrete antigens expressed by tissue or  
CC cell samples, and therefore used in humans for localization and  
CC monitoring of microbial infection  
XX  
SQ Sequence 871 AA;

Query Match 11.0%; Score 4533; DB 4; Length 871;  
Best Local Similarity 100.0%; Pred. No. 2.8e-226; Indels 0; Gaps 0;  
Matches 871; Conservative 0; Mismatches 0;  
QY 7098 MGPGLISLPGRKPGPCSSPGASQASSQVSSLRVGSQVGTTPGSLDAEGTQBAED 7157  
DB 1 MGPGLISLPGRKPGPCSSPGASQASSQVSSLRVGSQVGTTPGSLDAEGTQBAED 60  
QY 7158 LSDSTPTLQRFQEQATMRKFSLGGRGGYAGVAGYGTTFAGDAGGMLQGGPMWARIWAV 7217  
DB 61 LSDSTPTLQRFQEQATMRKFSLGGRGGYAGVAGYGTTFAGDAGGMLQGGPMWARIWAV 120  
QY 7218 SOSSEEEQEARAESQSEEQEARAESPLPOVSARPVPEVGRAPTRSSPEPTPWEDIGQV 7277  
DB 121 SOSSEEEQEARAESQSEEQEARAESPLPOVSARPVPEVGRAPTRSSPEPTPWEDIGQV 180  
QY 7278 SLVQIRDLSGDAEADTISLISEVDPAYNLSDLYDIKYLPPFFPMIPRKVPKSAQBPBP 7337  
DB 181 SLVQIRDLSGDAEADTISLISEVDPAYNLSDLYDIKYLPPFFPMIPRKVPKSAQBPBP 240  
QY 7338 SPMAEEELAEPEPTWMPGELGPHAGLEITEESDVEDDALLAEAVGKRKWSPSRSLF 7397  
DB 241 SPMAEEELAEPEPTWMPGELGPHAGLEITEESDVEDDALLAEAVGKRKWSPSRSLF 300  
QY 7398 HFPGRHLPDEPAELGLRERVKASVEHISRLIKRPEGLEKEGPPRKPGLASFRLSLGK 7457  
DB 301 HFPGRHLPDEPAELGLRERVKASVEHISRLIKRPEGLEKEGPPRKPGLASFRLSLGK 360  
QY 7458 SMDRAFTFLRELSDETIVLGOSVTLACQVSAQPAQATWSKDGAPLESSRVLISATLKN 7517  
DB 361 SMDRAFTFLRELSDETIVLGOSVTLACQVSAQPAQATWSKDGAPLESSRVLISATLKN 420  
QY 7518 FOLLITLVVVAEDLGVYTCVSNALGTVTTTGLVKRKAERPSSPCPDICEVYADGVLLVW 7577  
DB 421 FOLLITLVVVAEDLGVYTCVSNALGTVTTTGLVKRKAERPSSPCPDICEVYADGVLLVW 480

QY 7578 KPVESYGPVTYIVQCSLEGGSWTTILASIDFDCYLTSLKSRGTYTFTTACVSKAGMGPY 7637  
DB 481 KPVESYGPVTYIVQCSLEGGSWTTILASIDFDCYLTSLKSRGTYTFTTACVSKAGMGPY 540  
QY 7638 SSPSEQVLLGGPSHLASEEESQGRSAQPLPSTKTAFQTQIORGFVSVVQRCWEXASGRA 7697  
DB 541 SSPSEQVLLGGPSHLASEEESQGRSAQPLPSTKTAFQTQIORGFVSVVQRCWEXASGRA 600  
QY 7698 LAAKIIPYHPKDKTAVLREYEALKGLRHPHPLAQLHAAVLSPRHLVLIILELCSGPPELLPCL 7757  
DB 601 LAAKIIPYHPKDKTAVLREYEALKGLRHPHPLAQLHAAVLSPRHLVLIILELCSGPPELLPCL 660  
QY 7758 AERASYSESEVVDYLMQMSATQYLHNQHLHLDLRSENMIITEYNLLKVVDLGNQAUSLS 7817  
DB 661 AERASYSESEVVDYLMQMSATQYLHNQHLHLDLRSENMIITEYNLLKVVDLGNQAUSLS 720  
QY 7818 QEKVLPSDRFKDYLETMAPELLEGQGAVPQTDIWAIGVTAFIMLSAEYFVSSEGADLQOR 7877  
DB 721 QEKVLPSDRFKDYLETMAPELLEGQGAVPQTDIWAIGVTAFIMLSAEYFVSSEGADLQOR 780  
QY 7878 GLRKGVLVRLSRVAGLSGGAVAFILSTLCAQWGRPCASSCLOCPLWTEEGPACSRPAPV 7937  
DB 781 GLRKGVLVRLSRVAGLSGGAVAFILSTLCAQWGRPCASSCLOCPLWTEEGPACSRPAPV 840  
QY 7938 TPTETALRVFVRNREKRALLVKRHNLAQVR 7968  
DB 841 TPTETALRVFVRNREKRALLVKRHNLAQVR 871  
RESULT 13  
AAE16274  
ID AAE16274 standard; protein; 871 AA.  
XX AAE16274;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
XX Human kinase PKIN-20 protein.  
DE Human; kinase; PKIN-20; cancer; leukaemia; adenocarcinoma; osteoporosis;  
KW Immune disorder; atherosclerosis; Crohn's disease; Hodgkin's disease;  
KW Acquired Immune Deficiency Syndrome; AIDS; Addison's disease; anaemia;  
KW Allergy; asthma; adult respiratory distress syndrome; multiple sclerosis;  
KW autoimmune thyroiditis; bronchitis; diabetes mellitus; osteoarthritis;  
KW Good pasture's syndrome; Graves' disease; pancreatitis; psoriasis;  
KW rheumatoid arthritis; ulcerative colitis; cirrhosis; Cushing's syndrome;  
KW hepatitis; hypothyroidism; cerebral palsy; cataract; angina pectoris;  
KW cardiovascular disease; hypertension; vasculitis; myocarditis; obesity;  
KW congestive heart failure; ischaemic heart disease; lung tumour; gout;  
KW fatty liver; Niemann-Pick's disease; gene therapy.  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FT Domain 575..827  
FT /note= "Eukaryotic protein kinase domain"  
FT Domain 580..812  
FT /label= Protein\_kinase\_domain  
XX WO200196547-A2.  
XX  
PD 20-DEC-2001.  
XX  
PF 14-JUN-2001; 2001WO-US019444.  
XX  
PR 15-JUN-2000; 2000US-0212073P.  
PR 23-JUN-2000; 2000US-0213467P.  
PR 30-JUN-2000; 2000US-0215851P.  
PR 07-JUL-2000; 2000US-0216605P.  
PR 13-JUL-2000; 2000US-0218372P.  
PR 25-AUG-2000; 2000US-0228056P.  
XX (INCY-) INCYTE GENOMICS INC.  
PA

XX Yue H, Lal P, Bandman O, Borowsky ML, Au-Young J, Lu Y;  
 PI Gandhi AR, Tribouley CM, Wallia NK, Yao MG, Lu DM, Greenwald SR;  
 PI Ramkumar J, Griffin JA, Kearney L, Burford N, Nguyen DB, Tang YT;  
 PI Baughn MR, He A, Thornton M, Hafalia A, Patterson C, Gururajan R;  
 PI Lo TP, Khan F, Recipon SA, Azimzai Y, Pollick JL, Ding L;  
 PI Grether M, Elliott VS, Thangavelu K, Batra S, Ison CH;  
 XX WPI; 2002-090207/12.  
 DR N-PSDB; AAD26467.  
 XX  
 PT New polypeptides, useful for diagnosing, treating or preventing disorders  
 PT of growth and development, cardiovascular and lipid, and diseases such as  
 PT cancer, comprise human kinase polypeptides.  
 XX  
 PS Claim 1; Page 164-165; 197pp; English.  
 XX  
 CC The invention relates to human kinase PKIN proteins and their  
 CC corresponding cDNAs. A composition containing PKIN agonist is useful for  
 CC treating a disease or condition associated with decreased expression of  
 CC PKIN and a composition comprising PKIN antagonist is useful for treating  
 CC a disease or condition associated with overexpression of PKIN. The  
 CC disorders include cancer (leukaemia, adenocarcinoma, lymphoma, melanoma,  
 CC myeloma, sarcoma, teratocarcinoma, Hodgkin's disease); immune disorder  
 CC (Acquired Immune Deficiency Syndrome (AIDS), asthma, Addison's disease,  
 CC atherosclerosis, anaemia, allergies, adult respiratory distress syndrome,  
 CC autoimmune thyroiditis, gout, bronchitis, Crohn's disease, diabetes  
 CC mellitus, multiple sclerosis, Good pasture's syndrome, Graves' disease,  
 CC osteoarthritis, osteoporosis, pancreatitis, psoriasis, Reiter's syndrome,  
 CC rheumatoid arthritis, Sjogren's syndrome, uveitis, ulcerative colitis,  
 CC bacterial, parasitic, fungal, viral, protozoal and helminthic infections)  
 CC growth and development disorders (arteriosclerosis, cirrhosis, hepatitis,  
 CC Cushing's syndrome, hypothyroidism, cerebral palsy, cataracts); cardiac  
 CC vascular disease (arteriovenous fistula, hypertension, vasculitis,  
 CC aneurysms, congestive heart failure, angina pectoris, myocarditis,  
 CC ischaemic heart disease, chronic bronchitis, lung tumours); lipid  
 CC disorder (fatty liver, Fabry's disease, Niemann-Pick's disease,  
 CC hypcholesterolaemia, obesity). PKIN DNA is useful for assessing toxicity  
 CC of a test compound and in gene therapy. The present sequence is human  
 CC PKIN-20 protein  
 XX  
 SQ Sequence 871 AA;  
 Query Match 11.0%; Score 4529; DB 5; Length 871;  
 Best Local Similarity 99.9%; Pred. No. 4.5e-226;  
 Matches 870; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 7098 MGPDGSLPGRPKPGPCSSPGSSASQASSQVSLRVGSSQVGTPEPGLDAEGWTQEAD 7157  
 1 MGPDGSLPGRPKPGPCSSPGSSASQASSQVSLRVGSSQVGTPEPGLDAEGWTQEAD 60  
 7158 LSDSTPTLQRPQEQVTRKESLGRGGYAGVAGTGFAGDAGGMLGQGPWWARIANAV 7217  
 61 LSDSTPTLQRPQEQVTRKESLGRGGYAGVAGTGFAGDAGGMLGQGPWWARIANAV 120  
 7218 SQSEEEQEARASQSEEEQEARASPLQVARSVPVEVGRAPTRSSPEPTWEDIGQV 7277  
 121 SQSEEEQEARASQSEEEQEARASPLQVARSVPVEVGRAPTRSSPEPTWEDIGQV 180  
 7278 SLVQIRLSDGAEADTISLDSVDPAVNLSDLYDKLPFEFMRKVPKSAQPEPP 7337  
 181 SLVQIRLSDGAEADTISLDSVDPAVNLSDLYDKLPFEFMRKVPKSAQPEPP 240  
 7338 SPMAEELAFPEPTWMPGELPHAGLEITEESVDALLAEAAVGRKKNSSPSRSLF 7397  
 241 SPMAEELAFPEPTWMPGELPHAGLEITEESVDALLAEAAVGRKKNSSPSRSLF 300  
 7398 HFCRHLPLDEPAELGRERVKASVEHISRLKGRPEGLEKEGPPKKCLASFRISGLK 7457  
 301 HFCRHLPLDEPAELGRERVKASVEHISRLKGRPEGLEKEGPPKKCLASFRISGLK 360  
 7458 SWDRAPTFLRELSDETIVVLGQSVTLACQVSAQAAQATWSKDGAPLESSRVLISATLKN 7517

Db 361 SWDRAPTFLRELSDETIVVLGQSVTLACQVSAQAAQATWSKDGAPLESSRVLISATLKN 420  
 QY 7518 FOLLTILVVAEDLGVYTCVSNALGTVTTTGVLRKAERPSPPCPDIGEVYADGVLLAV 7577  
 Db 421 FOLLTILVVAEDLGVYTCVSNALGTVTTTGVLRKAERPSPPCPDIGEVYADGVLLAV 480  
 QY 7578 KEVESYGPVTYIVQCSLEGGSWTTLASDIFDCCYLTSKLSRGTVTFRTACVSKAGMGPY 7637  
 Db 481 KEVESYGPVTYIVQCSLEGGSWTTLASDIFDCCYLTSKLSRGTVTFRTACVSKAGMGPY 540  
 QY 7638 SPSQVLLGGPSHLASEESQGRSAQPLPSTKTFAPQTIQGRFVSVVRQWEKASGRA 7697  
 Db 541 SPSQVLLGGPSHLASEESQGRSAQPLPSTKTFAPQTIQGRFVSVVRQWEKASGRA 600  
 QY 7698 LAAKIIPYHPKDKTAVLREYEALKGLRHPHLAQHAAVLSPRHLVLILELCSGPELLPCL 7757  
 Db 601 LAAKIIPYHPKDKTAVLREYEALKGLRHPHLAQHAAVLSPRHLVLILELCSGPELLPCL 660  
 QY 7758 ABRASYSESEVKDYLMQWLSATQYLHNOHILHDLRSENMIITEYNLLKVVDLGNAGSLS 7817  
 Db 661 ABRASYSESEVKDYLMQWLSATQYLHNOHILHDLRSENMIITEYNLLKVVDLGNAGSLS 720  
 QY 7818 QEKVLPSDKFDYLETMAPELLEGQGVQPTDIWAIGVTAFLMSAEYFVSSEGARDLQ 7877  
 Db 721 QEKVLPSDKFDYLETMAPELLEGQGVQPTDIWAIGVTAFLMSAEYFVSSEGARDLQ 780  
 QY 7878 GLRKLGLVLSRCYAGLSGGAVALRSTLCAQPWGRPCASSCLOCPWLTTEGPACSRPAPV 7937  
 Db 781 GLRKLGLVLSRCYAGLSGGAVALRSTLCAQPWGRPCASSCLOCPWLTTEGPACSRPAPV 840  
 QY 7938 TPTTARLRFVVRNREKRALLYKRNHQAQR 7968  
 Db 841 TPTTARLRFVVRNREKRALLYKRNHQAQR 871  
 RESULT 14  
 ABG74786  
 ID ABG74786 standard; protein; 31267 AA.  
 XX AC ABG74786;  
 XX DT 05-JUN-2003 (first entry)  
 XX DE Human RGS11 protein.  
 XX KW RGS11; human; screening; cardiant; antianginal; gene therapy;  
 XX heart disorder; cardiac ischaemia; heart failure; angina.  
 XX OS Homo sapiens.  
 XX PN WO2002103355-A1.  
 XX PD 27-DEC-2002.  
 XX PF 17-JUN-2002; 2002WO-JP006019.  
 XX PR 18-JUN-2001; 2001JP-00183038.  
 XX PA (TAKE ) TAKEDA CHEM IND LTD.  
 XX PI Koyama N, Tanida S, Yamamoto K;  
 XX DR WPI; 2003-167557/16.  
 XX PT N-PSDB; ABX13540.  
 XX Screening compounds regulating RGS11 expression and activity for  
 XX prevention and treatment of heart disease.  
 XX PS Claim 1; Page 59-261; 321pp; Japanese.  
 XX This invention describes a novel method for screening compounds for their  
 CC ability to regulate the activity and expression of human RGS11 and its  
 CC partial peptides and salts, by observing the expression or activity of

CC RG511 in the presence or absence of the test compound. The products of  
CC the invention have cardiac and antianginal activity and can be used for  
CC gene therapy. The methods and compositions are useful in the prevention,  
CC treatment and diagnosis of heart disorders such as cardiac ischaemia,  
CC heart failure and angina. This sequence represents the human RGS11  
CC protein described in the disclosure of the invention  
XX  
SQ Sequence 31267 AA;  
Query Match 9.4%; Score 3878.5; DB 6; Length 31267;  
Best Local Similarity 21.2%; Pred. No. 4.6e-190;  
Matches 1954; Conservative 1220; Mismatches 3544; Indels 2493; Gaps 317;  
QY 10 PRFTRKAF-VVSGVGDATLSLCOVGNPTFQVSWKEDQDPVITAGARFLRAQDGLRLT 68  
DB 6861 PKFVKLEASKIVKAGDSRSLCKIAGSPRIRVVWFNEHELPA-----SDKYRMT 6911  
QY 69 ILD-----LALGDSGVVCCARNAIGBAFAAAGLQVDAEAAACAEQAPHELLRPTSI 119  
DB 6912 FIDSVAVIQMNLSTESGDFICQNPAGSTSCSTIV-----KEPVSFSPPIV 6964  
QY 120 KVRGSEATFRVGGSPRAVWSKGRRLGEPDGRVRVBEELGEASALRIARPRDG 179  
DB 6965 ETLKNAEVSLECELSGTPPFVFWVKQRQL--RSSKKYKIASKNFHTSIHLNVDTSDI 7022  
QY 180 GTVEVRAENPLGAASAAAL-----VVDSDAADTASRPGTSTAALLAHL 223  
DB 7023 GEYHCKAQNEVSDTCVCTVKLKEPPRFVSKNLSLTVAGEPABELQASIEGAQPIFQWL 7082  
QY 224 QRRREAMR-----AEGAPASPSTGTRTC----- 247  
DB 7083 KEKEEVIRESENIRITFVENVATLQFAKAEANAGKYICQKNDGMRENMATLWLEPA 7142  
QY 248 -----TVTEGKHARLSYVTEGPKPEVWKGDLQVTEGRRHVVEDAQENFVLKI 298  
DB 7143 VIVEKAGPMVTVTGETCTLECKVAGTPELSVWVKGDLTSQKH-KFSFYNKISLRI 7201  
QY 299 LFCQSDRLTCTASNLVGTYSVLLVVREPAVP--FKKRLQDLEVRKESATFLCEV 356  
DB 7202 LSVREQDAGTTFQVQNNVSKSCTAVVDSDRAVPPSFTRRLKNTGGVLGASCLCEKV 7261  
QY 357 PQPS-TEAANKBETRLWASAKYIEBEGTERRITVRNVSDADDVAVICETPE--GS--- 410  
DB 7262 AGSPISVAMFHEKTIKIVSGAKYQTTFSQNVCTLLQSLSDSDMGNYTCVAANVAGSDEC 7321  
QY 411 RTVAELAVQGNLRLKPRKTAVRVGDATMF-CVELAVPVGVMRLRQEEVVAGRVAIS 469  
DB 7322 RAVLTVOPPPFVKP-PEPLEVLPGKNTFTSVIRGTPPFKVNFRGARELVKGRCNLY 7380  
QY 470 AEGTRHTLTIQOCCLEDVQ-----VAFMAGDCQSTR----- 502  
DB 7381 FEDTVAELELEFNIDISQSGEYTCVSVNNAQASCTTRLFVKGLLKWIFFCFLYTDILVF 7440  
QY 503 -FCVSAKPPLOP-----PVDPVKARMESSVLWSPPHGERPVTI---DG 547  
DB 7441 FLCFLAVLCHLEPAAFKRLSDHSVEPKSIIILESTYT-----GTLPISVTWKKG 7492  
QY 548 YLVEKKLGTVTWIRCHAEAWATPELTADVABEENFQFRVSALNSFGQSPYLEFGTV 607  
DB 7493 FNITTS-----KCNIVTTEKTCILEILNSTKEDAGQYSCIEEAGRDVCGALVST- 7544  
QY 608 HIAPLKAVRTPKVAQAVEGVEVTSVLDLTVASAGE-----WFLDQALKASSVEIHCDR 663  
DB 7545 -LEPPYFV-TELEPLEAAVGSV--SLQCVAGTPEITVSWYKGDTKLRPTPEYRTFTN 7600  
QY 664 TRHTLTIREVPAASLHGAQLKFVANGI-----ESSIRMEVRAAPGLTANKPPAAAAREVLA 718  
DB 7601 NVATLVFNKVNINDSGEYTCKAENSIGTASSKTVEIOERQLPPSFAHQ-----LKDIEQ 7655  
QY 719 RLHFEAQLLAELSDQA-AAVWLKDGRTLSPGFKYEVOASAGRRVLLVRDVARDAAGLYE 777  
DB 7656 TVGLPVTLCRLNGAPIQVCWYRDGVLLRDDENLQTSFVDNVATLKILQTDLSHSGQYS 7715

QY 778 CVSRGRITAYQLSVQGLARFLHK-----DMAGSCVDVAVAGGPAQPECETSEAH-VHVHWYK 832  
DB 7716 CSASNPLGTASSABLTAREPKSPFFDIKPVSDIVIAGESADFECHVTGQAPRITWSK 7775  
QY 833 DQWELHSGSERPLOBVDVGRHRLVAATVTRQDEGTYSR-----VGEDSVDFLRLVSEPKV 888  
DB 7776 DNKEI-RPGNYTITCVGNTPELRLILKVGKSGSQYTCOATNDVGDKMCQAQLSVKEPKK 7834  
QY 889 VPAKQLARLKQAEAGASATLSCEVA-OAQTEVTWYKDGKLLSSSKVCKMEATCCTRL 947  
DB 7835 FVKLEASK--VAQGESIQLECKISGSPKLSWFRNDSHELHESWKNYNSFINSVALL 7891  
QY 948 VVQOAGQADAGEYSCEA-----GGQRLSFHLDVKEPKVWFAKDQVAHSEVQAEAGANATLS 1003  
DB 7892 TNEASAEDSGYICEAHNGVGDASCSTALTAKPVFTQKP-----SPVGLKGSQVILQ 7947  
QY 1004 CEVAQAQA-EVMWYKDGKLLSSSLKHVEAKCRRRLVVQOAGKTADAGYSCEARGQ--- 1059  
DB 7948 CEISGTPPEVWVWDRKQVRNKKFKITSKHFDTSHLINLEASDVGEYHCKATNEVGS 8007  
QY 1060 -RVSPFLHTITEPKWFAKEQSVHNEVQAEAGASAMLSCEVAQAQ-TEVTWYKDGKLL--- 1114  
DB 8008 DTCSCSVKPEPP-RPVKLS---DTSLTIGDAVELRAIVEGFQFISVVWLKDRGEVIRE 8063  
QY 1115 SSSSKYGMVEKCTRLRLVLPQAGKADAGEYSCEA-----EAGQQRVSPHLHITBPKGVFAKEQ 1170  
DB 8064 SENTRISFIDNIATLQGSPEA--SNSGKYICQIKNDAGMRECSAVLTVLEPARIEKPE 8121  
QY 1171 SVHNEVQAEAGTAMLSCEV-AQPQTEVTWYKDGKLLSSSKVMEVKCTRLRVVQVQV 1229  
DB 8122 ----PMTVTGNPFFALECVTGTPELSAKWFKDGRLELSADSKHITFINKVASLKIPCAE 8177  
QY 1230 KADAGEYSCE-----AGQQRVSPQLHITE---PKAVFAKEQLVHNEVTRBAGASATLSCEV 1282  
DB 8178 MSDKGLYSFEVKNVSKNCTVSVHVSDDRIVPSPFIRKLK---DVNAILGASVWLECRV 8233  
QY 1283 A-QAQTEVTWYKDGKLLSSSK-----VRTEAG---C 1311  
DB 8234 SGSAPISVGFODGNEIVSGPCQSSFSFENVCTLNLLEPSTDTGIYTCVAANVAGSDEC 8293  
QY 1312 MRQLVVQO-----AGQA----- 1323  
DB 8294 SAVLTVOEPPSPQTPDSVEVLPKMSLTFTSVIRGTPPFKVMFKSGRELVPGESCNISL 8353  
QY 1324 -----DAGEYTC-----EAGQRLSFHLDVSEPKAVFAKEQLAHKRVQAOEA 1364  
DB 8354 EDFVTELELFEVQPLESGDYSCLVNTDAGSASCTTHLFVKEP-ATFVK-RLADFSV--ET 8409  
QY 1365 GATATLSCE-VAQAQTEVTWYKDGKLLSSSKVMEAVGCTRLRVVQOACQADTGEYSCE- 1422  
DB 8410 GSPIVLEATYGTGTPISVSWIKDEYLTISOSECSITWTEKSTILELESTIEDIYAQYSL 8469  
QY 1423 ---EAGQRLSFSLDVAEPKWFAKEQPVHREYQAOAGASTTISCEV-AQAQTEVMWYKD 1478  
DB 8470 IENEAGQDICEALVSVLEPPYFI--EPL-EHVEAVIGSPATLQCKVDGTPEIRISWYKE 8525  
QY 1479 GKLLSFSSKVRMEAVGCTRLRVVQOAGQADAGEYSCEA---GSQRLSFHLHVAEPK--A 1532  
DB 8526 HTKLRSAPAYKMOFKNNVASLVINKVDSDVGEYSCKADNSVGVASSAVLVTKARLPP 8585  
QY 1533 VFAKEQPAQREVQAEAGTSATLSCEVAQAQ-TEVTWYKDGKLLSSSKVMEAVGCTRL 1591  
DB 8586 PFARK--LKDVTETLGFVAFECRINGSPELQVSWYKDGVLKDDANLQTSFVHNATL 8642  
QY 1592 VQOAGQADAGEYSCKA-----GDQRLSFHLHVAEPK-VFAKEQPAHREYQAOAGASATL 1646  
DB 8643 QILQTDQSHIGYNCASNPGLGTASSAKLILSEHEVPPFFDLKPV--SVDLALGSGTGF 8700  
QY 1647 SCEV-AQAQTEVTWYKDGKLLSSSKVMEAVGCTRLRVVQOAGQADAGEYSCE---AG 1701  
DB 8701 KCHVTGTAPKITWAKDNREIRPGNYKMTLVENTATLTVLVKVGKDGAGQYTCVASNIAG 8760  
QY 1702 GQRLSFHLHVAE-----LEPQISPRCPREPLVVKHEHDIILTATLTPSAAVTWNLK 1754



Db 8761 KDSCSAQLGVQPPRFIFKLLPSS-----RIVKQDEFTRVECKIGGSPKIVLWYK 8810  
 QY 1755 DQVEIRRRKHRTASQOHTLTHVGAOVLDSALYSCR-----VGAEGODFPVQVEEVAAK 1810  
 Db 8811 DETEQSSKFRMSFVDSVAVLEMHNLGVSDSGDYTCETHAAGSASSSTSLKVKKEPPI- 8869  
 QY 1811 FCRLEPPVCGELGGVTVLACEL--SPACAEVVRGCGNTOPRVRGKRFQWVAEGPVRSLTVL 1868  
 Db 8870 FRKKPHPIETLKGADVHLECELOQTPP-FHVSWYKDKRELSGKGYKIMSENFUTSIHIL 8928  
 QY 1869 GLRABDAGEYCESRDDHTSQAQLTVSV-----PRVKFMSGLSTVVAEBEGEATFOCVWS 1923  
 Db 8929 NVDAADIGEYQCKATNDVGSJTCVGSIALKAPPRFVKLSDISTV--GKEVQLQTTIE 8985  
 QY 1924 PSD-VAVVWFRD-GALLQSEKFAISQSGASHSLTISDLVLADAGQITVAEGASSS--- 1978  
 Db 8986 GAEPISVWVFMDKGEIVRESNIIWISYSENATIQGRVBPASAGKYTCQIKNDAGMQEC 9045  
 QY 1979 -AALRVREAPVLFKKKLEPQTVBERSSVLEVELT-RPWPEL--RWTRNATALAPGNVE 2034  
 Db 9046 FATLSVLEPATIVEK--PESIKVTGTCTLECTVAGTPELSTKWFKDKGKELTSDNKYK 9102  
 QY 2035 IHAEGARHRLVHNVGADRFQFCETPD-----DKTOAKLTVMRQV--RLVRGLQAVEA 2088  
 Db 9103 ISFNKVSGLKIINAVPSDSGVYFQVQNPVGKDSCTASLQVSDRTVPPSFTRKLKETNG 9162  
 QY 2089 REQQTATMEVQL-SHADVGSWTRDGLRFQOGPTCHLAVRGPMTHTLTLGLRPEDSG--- 2144  
 Db 9163 LSGSVVMECKVYSGPPIVSVMFHEGNEISSGRKYQITLTDNTCALTVNMBESDSGYT 9222  
 QY 2145 -LMVFKABGVHTSARLVVTELPVGSFRP-LQDVVTEKEKVTLECELSRPNVDVRLKOG 2202  
 Db 9223 CIA2NMAGSDCSAPLTVREPPSPFVQRPDMVLTGNVTFT-SIVKGTFFPFSVWFKGS 9281  
 QY 2203 VELRAGTMAIAAQCACRSLTIVRCEPADQGVVYCDADAQASASVKVQRTTYLIRRV 2262  
 Db 9282 SELVPGDRCNVLSDESVAELELFDVDTQSQGEYTC-----I 9317  
 QY 2263 LAEDAGIEQFVAENAEARQLRVKELPVTILVRPLRDKIANEMHRG-VLE-CQVSRASAQV 2320  
 Db 9318 VSNBAGK-----ASCETHLYIK-APAKFVKRLND-YSIEKGKPLILEGTFGTPIISV 9368  
 QY 2321 RWEKGSQELQPGKVELVDGLYERKLIISDVHAEDEDTYTC-----DAGDVKTSQAQFFVEE 2376  
 Db 9369 TWKNGINVTQRCNITITTEKSAILEIPSTVEDAGQYCNENASGKDSQCSAQIILIE 9428  
 QY 2377 QSI--TIVRGLQDVYVMEPAPAFECETSIPSVRPPKWLGLKTVLQAGNVG---LEQEG 2431  
 Db 9429 QKLPSPFSRQLRDVQETWGLPVFVFDCAI SGSEPIVSVMYKDGKPLKDSPNVQVSFLDNTA 9488  
 QY 2432 TVH-----RLMLRBTCTMTGPHFTVTKSRSARLVUS-----DIPVLTRELPBPK 2478  
 Db 9489 TLNIFKTRSLAGQYSCATNP-----IGSASSARLILTEGKNPPFPDILRLA---EVDVAV 9541  
 QY 2479 TGBELQSVLWSCDFR-----PAPKAVQWYKDDTPLSPSEKFKVSLGQMAELRILMPEA 2533  
 Db 9542 VGE-----SAUFECHVTGTQPIKVSMAKDSREIRSGGYQIYILENSAHLTVLVKDKG 9594  
 QY 2534 DAGYRQOA-----GSAHSSTEVTVAREV--TVTGPLQDAEATEEGWASFSCE---LSHE 2584  
 Db 9595 DSGQYTCYAVNEVGKDSCTAQLMIKERLIPPSFTKRLSETVBETEGR-NSFKLEGRVAGSQ 9653  
 QY 2585 DEEVESLNGMPLYNDSPHISHKRRHTLVLSKIORA---DAGIVR-----ASSLKVS 2635  
 Db 9654 PIIVAWYKNNIEQPTSNCEITFK--NNTLVLO-VKAGWMDAGLYTCVKNVNDAGSALCT 9710  
 QY 2636 TSARLEVRVPVFLKALDLSAEBEGTLALQCEVSDPE-AHVVVRKQGVQIGPSDKYDF 2694  
 Db 9711 SSVIKPEKKPPFPDQHLTFTVTVSEGYVQLSCHVQGSFPIRQIOMLAKGREINKPSDRCSF 9770  
 QY 2695 LHATGRGLVVDVSPEDAGLYTCH---VGSSETRARVRVHDLHVGTIKRLKTWEVLBG 2750

Db 9771 SPASGTAVLELDVAKADSGDYVCKASVAGSDTTKTSKVITKAEPQITKRIQNIWSEH 9830  
 QY 2751 ESGSEFVLSHESASDPAPWTVGGKTVGSSSRFQATROGRKYILVVRBAAPSDAGEVVS 2810  
 Db 9831 QSATECEVSPDDAI--VTWYKGPTELTSQKYNPRNDGRCHYMTIHNVTDPDEG--VYS 9886  
 QY 2811 V-----RG-----LTSK-----ASLIVR----- 2823  
 Db 9887 VIARLEPRGEARSTAEYLITTKIEIKLELKPDPIDPSRVPIPTMPPIRAVPPBEIPVAPP 9946  
 QY 2824 -----ER-----PAALIKP-----LEDQ 2836  
 Db 9947 IPLLPTPEKKPPKRIEESHKRVPAKVPKAPPKPKVVKPVIKIEKTSRMEEE 10006  
 QY 2837 WV-----AP----- 2840  
 Db 10007 KVQVTKVPEIKPAIPAPPEPKPEAGVPKTPSPIEARRKLRPGSGGKPPDEAPFT 10066  
 QY 2841 -----GEDVELRCELSRAGTVPWMLKDRKAIRKSQYDV 2875  
 Db 10067 YOLKAVPLKFVKEIKDIILTESEFVGSASAIPECLVSPSTAITTWKOGSNIRESPKHPFI 10126  
 QY 2876 CEGTWAMLVIRGASLKDAGEYTCV-----EASKSTASLHVEBKANCFTTEELTNQVEBK 2931  
 Db 10127 ADGDKDKLHIIDVQLSDAGEYTCVLRNLGNKEKTSKALWBEELPVRFVKTLEEVTVVKG 10186  
 QY 2932 TAVFTCKTEHPAATVTRK-GLLELRASGKHOPQEGTLTLTISALEKADSDTYTCDIG 2990  
 Db 10187 OPLYLSCELNKRDVVRKDKIVVEKGRVPGVIGLVRALTIINDADDDTADGTYTIVE 10246  
 QY 2991 QAQ-----SRAQLLVGRRVHIIEDLVDVQEGSSATFRCSIPANPEPVHWF--LDKTP 3044  
 Db 10247 NANNLESCVKKVVEIRDVLWKPTRDQHVFKGTAIFACDIA-KOTPNIKFKGYDEIP 10305  
 QY 3045 LHANELNIIDQAQGGYHVLTLRQALAKDSGTIYEAGDQASAAALRVTEKPSVSRDLTD 3104  
 Db 10306 AEPNDKTEI-IRDGNHLYLKIKNAMPEDIAEYAVEIEGKRYPAKLTIGEREVELLKPIED 10364  
 QY 3105 ATITEGEDLTVCETSTCDIPMCWTQDKTLRGSARQLSHEGHRALQTLITGATLQDSGR 3164  
 Db 10365 VTIYKESASPDABISEADI PQGKMLKCELLRPSPTCEIKAEKGKFLTLHKVLDQAGE 10424  
 QY 3165 YKCEAGGACSSIVRVHARPVRFQEBALKOLEVLGGATLRCVLSSVAAAPVKWCYGNVL 3224  
 Db 10425 VLYQALNAITTAITLVKBIELDFAVPLKDVTPERRQARFECVLTR-EANVMSKGPDI 10483  
 QY 3225 RPKGYSLRQEGAMLELVVRNLRPQDSGRYSCSFGDQTTSATLTATVTAQAFIKLANKE 3284  
 Db 10484 KSSDKFDIADGKHILVINDSQDDEGVYTAEEVGKTSARLFTVGTIRLKFNSPLEDQT 10543  
 QY 3285 ATEGATATLRCELS-KTAPVWRKGSSETLRDGYCYLRQDGMCELOIRGLAMVDAEYS 3343  
 Db 10544 VKEGETATFVCELSHEKXHVVMFKNDKALHTSRVTLISSEKTKHLEMKVETLDDI SQIK 10603  
 QY 3344 CVCGEERTSALUTRPMFAHPIGRRLRHOESTEGATATLRCELSKAAAPVWRKGRESURDG 3403  
 Db 10604 AQVELSSTAQLKVLADPPTVTKLHDKTAVEKDEITLKEVSKDVVPVKFQOGEIUPS 10663  
 QY 3404 DRHSLRDGAVCELOICGLAVADAGEYSVCVGEERTSATLTIVKALPAKFTFEGELRNEAVE 3463  
 Db 10664 PKYSIKADGLRILKIKADLKDKGEYVDCGDTKTKANVTVEARLIKVEKPLYGVFV 10723  
 QY 3464 GATAMLWCELS-----KVAPVWRKGPENLRDGYRILRQEGTRCELOICGLAMA 3513  
 Db 10724 GETAHPFIEISEPPVHGWKLKQPLTASPCDEIIEGKHIL-----ILHNCQLGM- 10775  
 QY 3514 DAGYLCVCGOERTSATLTITRALPARFIEDVKNQAEAGATVLAQCELN--SAAFVWRK 3572  
 Db 10776 -TGEVSPQAANAASAAALKVKELPLIFITPLSDVVKPEKDEAKFECEVSRPKTFRMLKG 10834  
 QY 3573 SETLRDGRYSLRDGTKCELOIRGLAMADTGEYSCVCGOERTSAMLTVRALPIKTEGL 3632  
 Db 10835 TQBITGDRELKDGFKHSHWIKSAFADEAKFMEADKHTSGKLIIEGILKELTJPL 10894



QY 3633 RNEEGATEGAV----- 3644  
D 10895 KDVTAKESAVFVELSHDNIRVKWFKNDQRLHTRSVQMDBGKTHSITFKDLSIDT 10954  
QY 3645 -----LRCELSM-APVEMWKGHE 3662  
D 10955 SQIRVEAMGMSSEAKLTVEGDPYFTGKLQDYTGVEKDEVILQCEISAKADAPVKWFKDGK 11014  
QY 3663 TLRDGRHSRLQDARGCELOIRGLVAEDAGBYLQWCKGERTSAMLTVRAMPSEFIEGLRN 3722  
D 11015 EIKSKNAVIRKADGKRMILUKKALKSDIGQYTCDCGTDKTSGLKDIEDREIKLVRPLHS 11074  
QY 3723 BE----- 3724  
D 11075 VEVMETETARPETBEISDDIHANWKLKGEALLQTPDCEIKEBEGKIHSLVHLNCRLDQTGG 11134  
QY 3725 -----ATGSDATILWCELS-KAAPVEMWKGHE 3752  
D 11135 VDFQANVKSAAHURVKPRVIGLRLPLKDVTVTAGEATFOCELSYEDIPVEMWKGKGL 11194  
QY 3753 RDGRHSRLQDGRSCELOIRGLVAEDAGBYLQWCKGERTSAMLTVRAMPSEFIEGLRN 3812  
D 11195 BPSDAVVRSEGVHTLRLDVKLEDAQEVQLTAKDFKTHANLVKBPVPEFTKPLEDQT 11254  
QY 3813 ABEGATAVLOCELSKA-APVEMWKGSETLRGDRYSRLQDGTGRCELOIRGLSVADTGEYS 3871  
D 11255 VEEGATAVLECEVSRENAKWFKNGTEILSKYIEIVADGRVKLVHDCPTEDIKTYT 11314  
QY 3872 CVCGQERTSATITVRAPQVFPREPQLQSQABEGSTATILQCELSBPTATVVMWKGGLQLOA 3931  
D 11315 CDKDFKTSNVLNVPVPHVFLPLTDLQVREKEMARFECELSRENKVKWFKDGAIEKK 11374  
QY 3932 NGRREPRLOGCTAEVLVLODQREDTGEYTCGSGQATSATITVTAAPVFLRELHQEVD 3991  
D 11375 GKXYDIISKGAVRILVINKCLLDDAEYSCEVARTARTSGMLTVEEAVFTKNLANIEVS 11434  
QY 3992 EGGAHLCELSRAGASVEMWKGSLQLPFCARYQWVODGAAALVRGVEQEDAGDVTCD 4051  
D 11435 ETDTTLKVEKSGPQREWINYKGEDEIETGRYEILTEGRKILVQNAHLEDAGNYCR 11494  
QY 4052 TGHQSMASLSVRPRPKTRQLQSLQEBQTDIARLCCQLSDASGAVVQWKGVELHA 4111  
D 11495 LPSRTDGVKVKHELAEFISKPNLEILEGEKAEFVCSIS--KESFPVQWKRDDKILES 11552  
QY 4112 GKPYEMRSGATRELLIHQLEAKDTGEYACTVGGOKTAASLRVTEPEVTIVRGLVDAEVT 4171  
D 11553 GDKYDVIADGKRVILVWKDQTLQDMGTVMVWGAARAAHLTVIE-KURIVVPLKDRVK 11611  
QY 4172 ADEDEVFCSVRAGATGVQWCLQGLPLQSNVEVAVRQRIHTRLRKGVTPEDAGTVS 4231  
D 11612 EQQEVVFNCEVNTGAKAKWFRNEEAI PDSKY--IILQKDLVYTLIRDAHLDDQANYN 11669  
QY 4232 FHLGNH-----ASSAQLTVRAPEVITILEPLQDVQLSEGDASFCQLRSRSGQEARWALG 4286  
D 11670 VSLTHRGENVKSAANLIVEEDRLRIVEPLKDIETMEKSVTFWCKVNRNL-VTLKWTKN 11728  
QY 4287 GYPLQANEMNDITVEGTLHLTLKVKLTLEDAGTVSFHVGTCSSEAQLK-VTAKNTVVRG 4345  
D 11729 GEVFPDNRVSRVK-KYHMLTIDKCGFPDEGEIIVTAGQDKSVAELLITIEAPTEFVEH 11787  
QY 4346 LENVEALEGEALFECQLSQPEVAATHLLDDDEPVRTSENARVFFENGLRHLLLNLR 4405  
D 11788 LEDQVTPEPDDAVFSQLSR-EKANWYVRNGREIK--EGKKYKPEKDGSIHRIIICDR 11844  
QY 4406 PQDSCEVTLFAGDWTYSALTVGRWRLTEILEPLKNAVRAGAAQAFCTCLSEAVPVGEAS 4465  
D 11845 LDDECBYACGVEDRSARLFFVEEIPVEIIRPPQDILEAPGADVDFLAEIKN--DKVEVQ 11902  
QY 4466 WINGAAVQDDSDMTVTADGSHQALLRSQAQFHAGEVTFACRDVAASAR----- 4516  
D 11903 WLRNNMVVQGDKH-QMMSEGIHRLIQICDIPRQDQGEVRFIAKDKEARAKLELAAPKI 11961  
QY 4517 ----- 4516  
D 11962 KTADQDLVVDGKPLTMVVPYDAYPKAEAEWFKENEPISTKTIDTABBQTSRILEAKKG 12021  
QY 4517 -----LTVLGLPDPPEDAEVAWSSHTVTLSSWAAWPMDSGGGGLC 4555  
D 12022 DKGRYKIVLQNKHGAEGFINLKVLDVPCPVNLEVTTFDGEVSLWAMEEPLTDGSGKII 12081  
QY 4556 GYRVEVKEGATGWLCHLHELVPGBECVVDGLAPGET-YRFRVAAGVPVAGBPHVLPQTV 4614  
D 12082 GYVVERDRIKRTWILATDRAESCEFTVTGLOKGGVEYLFVRVARSARNVGTGPBVTDNVP 12141  
QY 4615 RLAEPPKPVPPQ-----SAPESRQVAAGEDVLELEVAEAGEVHWHKM 4660  
D 12142 E-ARSKYDVPGLPLNTITDNRFGVSLTWEPEDVGGAEITNYVIELDKTSIRWDTAM 12200  
QY 4661 ERIQGRPEVYSQGRQQLVIGFTAEQDGBYHCGLAQSGICPAAATFQVALSPASVDE 4720  
D 12201 T-----VRAEDLSATVTDVVEQGEYSFRVQNRNIGVGK-----PSAATPFVADPIER 12251  
QY 4721 APOPSLPPEAAQEGDLHLWEALAKRMSRPTLDSISELPEEGRSQRLOPQAEVAP 4780  
D 12252 SPVNLITSDQTSQSSVOLKWEPLKD---GGSPILYIIIERCEE-GKONWIKNM-KLVP 12306  
QY 4781 DLSEGYSTADELARTGDADLSHTSSDDESRAGTPELTYLKAGRGFTSPLASKVGAPAA 4840  
D 12307 ELTYKVTGLEK---GNKYLVRVSAENKAGVSDPSEI----- 12339  
QY 4841 PSVKPQQOQEPPLAAVRPPLGLDLSKTD-LGDPMSDKAAVK-----TCAAFKGYK 4887  
D 12340 -----LGPLTADDAFVEPTMDLSAFKXGDEVIPNITILVPTGY- 12380  
QY 4888 VRKEMKQGBPMPSHTFGTEAQVGDALRLECVASKADVRARWLKGDVELTDGRHHHD 4947  
D 12381 -----PRPATWCFGDKVLETGDRVKMLISA----- 12407  
QY 4948 QLQDGTCSLLIAGLRADAGCYTCQVSN-----RFGQVTHSA 4984  
D 12408 -----YAEVLISPSERSDXGIVTLKLENVKITSGEIDVNVIAIARPSAPKELKFGDITKDS 12462  
QY 4985 CVTVSGSEAESESSSGELDDAPRAARLRLFRTK----- 5021  
D 12463 VHLT-----WEPDDDGSPITGVVKEKREVRKTKWMDVTDLEFTVPDLVQKEYLF 12518  
QY 5022 -----SPARYSDEELFLSADGPAEPEPADWQTYREDEHFI----- 5058  
D 12519 KVCARNKCGPGEPAVY-DEPVNMSTPATVPDPENVKWRDRTANSIFLTWDPKPPKDGSR 12577  
QY 5059 -----CIR-----FEALTARQAVTRQE--MEATLIGIVEIKLVEQGPBRVEM 5100  
D 12578 IKGYIVERCPRGSKWVACGEPAETKMEVTGLEGKWA-----YRVKA 12622  
QY 5101 CISKETAPVVPPEPLPSLLTSDAAPVFLTE--LQNOEVQDGYPSFDCVVTQCPMPSVR 5158  
D 12623 LNRQGASKPSRPTTEECAVDTOEAPIFLDVKLLAGLTVKAGTKIELPAITVTKPBPKIT 12682  
QY 5159 WFKDGLLEBDDHYMINEDQGGHQIITIAVVPADMGVTRCLAENSGVSTKAEIRV-- 5216  
D 12683 WTKADMILKQDKRITI-ENVPKSTVTIVDSKESDTGTVIIIEAVNCGRATAVVENVLD 12741  
QY 5217 -----DLTS-----TDYDTAADATESSEYFSAQCYLSSREQ 5247  
D 12742 KPQPPAFAFDITDVNESCCLTNWPPRDDGSKITNVVERATDSEVWHKLSSTVKDNTF 12801  
QY 5248 EGTESTTDEQLPQWVEEL-----RLQVAPGTRLAKFOLKVKGY- 5291  
D 12802 KATKLIENKEVIFRVAENNMVGEVPQASPIIT--AKYQDPDPGPPTRELPSPDITKDAVT 12859  
QY 5292 LYWFK-----DQOPLTASAHIRM--TGKKIL-----HTLEIISVTREDSGYAAIYS 5336  
D 12860 LTVCEPDDDGSPITGVWVERLDDTDKVRCKNMFKVOTTYRVKGLTNKKKRVFRVLA 12919  
QY 5337 NAMG---AAYSARLLVRGPEPEEKPSADVHEQLVPPRML-----RFTPKV 5382

Db 12920 NLAGPKFKSPSTEPILIKDIPDPS-----PPRWLEVINITKNTADLKWTVEK 12968  
 QY 5383 KKGSSITTSVKVEGRPVPTVW-----LREEAERGVLNIGP 5418  
 Db 12969 DGGSPTNVI--VEKDVRRKGQVDTTVKTKCTVPTPLTEGSLVYFRVAENA---IGQ 13024  
 QY 5419 D-----TPGTVASSAQOHSILVLLDVGRHQGTGTYTCTIASNAQAALCSA 5462  
 Db 13025 SDYTEIEDSVLAKDTFTTPG-----PPYALAVVDVTKRHVDLKWBPXKNGDGRPI--- 13074  
 QY 5463 SLHVSGLPKVEBOE-----KVKEALISTFLOGTTOAISAOGL----- 5499  
 Db 13075 QRYV-----IEKRELGRWVKAGTAGPCNCFRVDVIEGTVEQFQVRAENEAGVGHPS 13129  
 QY 5500 ---ETASPAD-----LGGQRKEEPLAAK--EALG-----HLSLAEVGTGE 5534  
 Db 13130 EPTETLSIEDPTSPSPPLDLHVTDAGRKHTAIANKPEKNGGSPILGYHVEMCPVTEK 13189  
 QY 5535 FLO-----KLTQITEMVSAKITQAK----- 5555  
 Db 13190 WMRVNSRPIKOLKFKVEGVVDPKEYVYLRVAVNAIGVSESEISENVVAKDPCKPTID 13249  
 QY 5556 -----LQVPGGSDSDSKTSPASPRHGRSRSPSSIOESSSE 5593  
 Db 13250 LETHDIIIEGKLSIPVPRFRAVPVTVSWHKQKEVKASDRLTMKNDHISAHLEVPKSV 13309  
 QY 5594 DGDARGETFDIYVVTADYPLGABQDAITLREGQYVEVLDAAHPLRWLVRTKTKSPSR 5653  
 Db 13310 RADA-----GIYTTILEN--KLGSATASIN-----VKVIGLPGCKDIKASDITKSS--- 13354  
 QY 5654 QGWTSPAYLDRRLKLSPEWGAABEPPEGEAVSEDEYKARLSVIOELLSEQAFAVEELQ 5713  
 Db 13355 -----CKLT--W--EPPEFDGG-----TPILHYVLEERRA----- 13380  
 QY 5714 FLOSHHLOHLRCPHPIAVAGQKAVIFRNVRDIDIGRFHSSFLQBLQOCDTDDVAMCFIK 5773  
 Db 13381 -----GRTYVIPV--MSGENKLSW--TVKDILPNGEYFFR-----VK 13413  
 QY 5774 --NOAAFEQYLEFLVGRVQABSVVSTAIQEFYKYABEALLAGDPSQPPPPLOQLVLEQ 5831  
 Db 13414 AVNKVGGEYIEL-----KNFVIAQDPKQPPDPVDEVEHN 13449  
 QY 5832 PVERVQRYQALLKELIRNARNRQNCALLQEAQVAVSALFORAENKLHVSLMENYPTGLE 5891  
 Db 13450 PT-----AEAMTITWKPL-----YDGGSK 13469  
 QY 5892 ALGEPHQGHIVWEGAPGAPMPKGNHNVFLFRNHLVICKPRRSRTDTVSYFRNMW 5951  
 Db 13470 IMG-----YIIEKIAGEER--WRCNEHL-----VPILTYTAKGL- 13503  
 QY 5952 KLSSIDLNDQVEGDRDAFVWQEREDSVRKYLLQARTAIKSSWKEICIGIQORLALPVW 6011  
 Db 13504 -----BEGKEYQFRAENAAGIS-----EPSRATPPT 13531  
 QY 6012 R---PPFEBELACTAEL--GETVKLACRVGTQPKFVSWYKDGKAVQVDPHHILIEDP 6066  
 Db 13532 XAVDFIDAPKVILRTSLVLRGDEIALDASIGSPYTTIWKDENVI----- 13579  
 QY 6067 DGSCALILDSITGVDVSGQYMCFAASAAGNCSTLGLKILVQVPPFRVNVKVRASPRVEGEDAQ 6126  
 Db 13580 -----VPEEI--KXGAAPLVRRRKGE 13598  
 QY 6127 FTCTIEGAPYQIRWYKDGALLTGNKFQTLSEFRSGLLVLVIRAAKSEDLGYECELVN 6186  
 Db 13599 VQ-----EBEPF-----VLPLTQRL--SIDNSKKGESQLRVRDLSLRPDHGLYMKIVEN 13644  
 QY 6187 RLQSARASABRLTQS-----PMLQAOEQCHREQLVAVE-----DTILERADQB- 6230  
 Db 13645 DHGIAPACTVSLVDTGPGPINFVFDIRKTSVLCKWEPLDDGGSEIINYILEKKDKTK 13704  
 QY 6231 -----VTSVLK-----RLIGPK-----APGPS-----TGDLTGPGPC 6257

Db 13705 POSEMIWVTSTLRHCKYSVTKLIEGKEYLFRVRAENRFGPGPPCVSKPLVAKDPFGPPDA 13764  
 QY 6258 P-----RGAPAL-----QETG----- 6268  
 Db 13765 PDKPIVEDVTSNMLVKMNEPKDNGSPILGYWLEKREVNSTHWSRVNKSLLNALKANVDG 13824  
 QY 6269 -----SQPPVTGTSAPAVPRVPQPLLHLEGPEOEPEAIARAQ 6306  
 Db 13825 LLEGLTVYFRVCAENAAAGPKESPSPDKTAHDPISPGPPPIPRVYTDI--SSITIEL----- 13879  
 QY 6307 EWTVPIMEGAWPG-----AGTCELLWD-----VHSVVRRETQ--RTTYQOAI 6350  
 Db 13880 EWEPPAFNGGGEIVGYFVDKQLVGTNE--WSRCTERMIKVRQYTVKEIREGADYKLRVSA 13937  
 QY 6351 THTA-----RPPSMQVTIE--DVQAQTGGTAQOFAELIEGDPOPSVTW 6390  
 Db 13938 VNAAGEGPPGETQPVTVABEQPFAVELDVSKGIGIOMAGKTLRIPAVVTRGPVPTKVW 13997  
 QY 6391 YKDSVOLVDSTRLSQOQEGTYSVLVRHVASKDAGVYTCLAQNTGGQVLCKAELLVLGD 6450  
 Db 13998 TKEEGEL--DKDRVWIDNVGTKSELI--KDALRKDHGRYVITATNSCGSKFAAARVEVF--- 14053  
 QY 6451 NEP-----DSEKQSHRRK--LHSFYEVKEEIGGVFGVFRVQHKGNKILCAAKFTPL--RS 6503  
 Db 14054 DVPGPVLLDKPVVTVNRKMLNWSDEDDGGEITGFI--IERKDAKVHTWRQPIETERS 14111  
 QY 6504 R-----TRAQYRER-----DILAA--LSHPLVTGLLDOFE--TRKTLI 6538  
 Db 14112 KCDITGLLEGQYKERVAKNGKFGCPPEIPIILAVDPLGPTSPESLTVTERTKSIIT 14171  
 QY 6539 L-----ILE-----LCSSELLDLRLKGVVTVAEVKVY 6567  
 Db 14172 LMKKEPRNGSGSPIQGYIIEKRHDKDPFERVNRKLCPTTSFLVENLDEHOMYFRVKA- 14230  
 QY 6568 IOQLVEGLHYLHSHGVHLHDIKPNSILMVHPARDEIKICDFGAQNTIPABLOFSQYQSP 6627  
 Db 14231 VNEIGESPSPLNVVIOQDEVPPTIKLRLSVRGDTIKVKAGEPVI--PADVTGLPMPKI 14289  
 QY 6628 ERFVSFEITQOPVSPASDIWANGVI--SVLSITCSPPAGESDRATLLNVLEGR----- 6679  
 Db 14290 EWSKNETVIEKP--TDALQITKEEVSRSSEAKTELSIPKAVREDKGTVYTVASNRLSGVFRN 14348  
 QY 6680 -----VSWSSPMAAHLSEDAKDFIKATLQAPQAPSA 6713  
 Db 14349 VHVVEYVDRPSPRNLAVTIDIKAESCYLTWAP-----LDNGSEITHVIVDKRDSRKA- 14403  
 QY 6714 QCLSHWFLKSNPABEAHF-----INTKQLFLLRGRWQSLMSY-----KSILVNR 6762  
 Db 14404 -----EWBEVNTAVEKYGIWKLIPINGQYEFVR-----RANKYGISDECKSKVVIQ 14452  
 QY 6763 IPELLRGPPDPSLGVARHLCRDTCGSS--SSSSSSSNELAPFA-----RAKSLPP--SP 6813  
 Db 14453 DPYRLPGPGPKV--LAR-----TKGMLVSWTVPDLNGGSPITGYWLEKREESPTWSR 14506  
 QY 6814 VTHSPL-----LHPRGFLRPSASLPEEAESERSTEAP 6846  
 Db 14507 VSRAPITKVLKGVFEFNVPRLEGVKYQFRAMANAAGIPGPPSE--PSDPEVAGDPIFP 14564  
 QY 6847 APPASPE-----GAGPRAAQCVPHSVIRSLFVHQAGESPEHAGLAFGRSRHP 6895  
 Db 14565 GPPSCPEVKDKTSISISLUGWKPPKADGGSPIKGYI-----VEMOEGTDTWKRVNPE 14616  
 QY 6896 ARRRHLKGGYIAGALPCLREPLMEHRV--LEBEAAREEQATLLAKAPSPETALRLPAS 6952  
 Db 14617 DKLITTCB-----CVVFNLKE--LRKYFRVKAVNEAGESE-----PS--DITGEIPAT 14661  
 QY 6953 GTHLAP-----GHSLSLEHDSPTSPPSSEAC--GEAQLPSAPSGAIFRDMGHQGS 7004  
 Db 14662 DIOBEPEVFDIGAQDCL-----VCKAGSQIRIPAVIKO----- 14695  
 QY 7005 KQLPSTGCHPGHTAOPERPSPDPSPMQPPAPFCHPKQGSAPQEGGSPHVAFCPPGSPPG 7064  
 Db 14696 -----RPTPKSSW----- 14703

WPI; 2001-451869/48.  
 N-PSDE; AAS05390.  
 Determining if a subject has or is at risk of developing a titin-related disease or condition, particularly heart failures, comprises detecting the presence of a mutation in the titin gene.  
 Disclosure; Page 57-111; 114pp; English.  
 The present sequence representing human titin (also known as connectin) is described in an invention relating to a novel method for determining whether a subject has or is at risk of developing a titin-related disease or condition. The method comprises analysing a nucleic acid of a sample from the subject and detecting the presence of a mutation (e.g. the pickwick mutation in the cardiac specific exon N2B) in the titin gene, which indicates that the subject has or is at risk of developing a titin-related disease. The zebrafish which has a phenotype similar to mammalian heart failure is used as a model. The method is useful for detecting an increased likelihood of heart disease, such as heart failure, in a patient, so that appropriate intervention can be instituted before any symptoms occur. The method may also be used to facilitate determination of etiology of an existing heart condition, such as heart failure, to identify compounds that can be used to treat or prevent heart conditions, in prenatal genetic screening, e.g. to identify parents who may be carriers of a recessive titin mutation. Compounds identified using the methods may be used to treat patients that have or are at risk of developing heart disease, e.g. heart failure.  
 Sequence 26926 AA;  
 Query Match 8.7%; Score 3580.5; DB 4; Length 26926;  
 Best Local Similarity 20.2%; Pred. No. 1.1e-174;  
 Matches 1930; Conservative 1288; Mismatches 3661; Indels 2663; Gaps 320;  
 QY 4 PQFGSARFRTRPKAFVVSVGKATLSQVLGNPTPOVSNEKQQPVTAGARFLA---Q 60  
 DB 1030 PGEPAAPYFITKPVQKLVGGSVFGQVGGNPKPHVYKKSQVLTGTRYKVSYNKQ 1089  
 QY 61 DGDLYRLTILDALGDSQVYVCARNAIGRAFAAVGL-----QVD 100  
 DB 1090 TGEC-KLVISMTFADDAGEYIVVRNKHGETSASASLLEADYELLMKSQOQMLYQTQVT 1148  
 QY 101 A---EACACGQAPHFL----- 113  
 DB 1149 AFVQEPVGETAPGVFVSEYKEKEQALIRKKOMAKTVVVRVYVEDQEFHISSFERL 1208  
 QY 114 -----LRPTSIRVRGSGSEATPRCV 133  
 DB 1209 IKETRYRIKTTLEELLEDEGEKQAVDISESEAVESGFDLRINKYRLGSMGVTFCHKM 1268  
 QY 134 GGSPPRAVSWSKGRRLEGDPGRVVRVEELGEASA-LRIRAAPRDPGGTVEVRAENPLGA 192  
 DB 1269 SGYPLPKIANYKDGKRI--KHGEYQWDFLDGCRASLIRPVLPEDSGIYTAFAASNKGN 1326  
 QY 193 ASAAAAVVDSDAADTASR--PGTSTAALLAHLQRR--REAMRAEGAPASPP----- 240  
 DB 1327 AICSGKLYVEPAAPLGAFTYPTLEPVSRIETSLSPRVSRSPIRMSPARMSPARMSPARM 1386  
 QY 241 -----STGTR-----TCTVTTEGKHAULSCYVTEGPKETVWVK 273  
 DB 1387 SPARMSPGRRLEETDESOLERLYKPVFVKPVSFKLEGQATRDLDKVGVRMPDETWFH 1448  
 QY 274 DGQLVTEGRRH--VYVEDAQENFVLKILFCKQSDRGLYTCTASNLAVQGTYSVSLVAVV--- 328  
 DB 1447 DGQIQVNDYTHKVYKEDGTQSLI--IVPATPSDGSGETVVAQNRAGRSSISVLITVEAV 1504  
 QY 329 ----- 328  
 DB 1505 EHQQKPFVEKIKVNIKEGSRLEMKVYRATGNPNPDIVLWLNKSDIIIVPHKYPKIRIEGK 1564  
 QY 329 -----R 329  
 DB 1565 GEAALKIDSTVSQDSAWYTATINAKGRDITTRCKVNVEVEFAEPEPEKLLIPRGTVRAK 1624

QY	7065	SCKEAFLVPSSTFLGQPAPAPAKASPLDSKMGDISLP--GRPKPGCSFSGASQ	7122
Db	14704	-----BFDGK-----AKAMKLETAENSSVIIPECKRSHTGYSI--TAKN	14743
QY	7123	ASSOVSRLRVSSGVGTPEGPSLDAEGWTOEAELSDSTPLQRPQQATMRXFSLGG	7182
Db	14744	KAQCKTANCRV--KWDVPGPDKLK-----VSDIT-----R	14773
QY	7183	GGYAGVAGYTFAFGDAGCMLGCGPMARIWAYSQSSEEEQEAEARASOSEEQEARA	7242
Db	14774	GS-----CRLSWKMPDDDDGGDIKGYY-----	14795
QY	7243	ESPLPOVSAREPVEVCRASTRSPPTPWEDIGQVSLYQIRDLSCDAE-----	7290
Db	14796	-----IEKRITID--GKAWTKNP-----DCGSTTFV-VPDLLSEQQYFRVRAENRF	14840
QY	7291	--AADTISLDISEVDPAYLNLSLDYIKYPFEFMIFRKVPSAQPPPPSPMAEEELAE	7347
Db	14841	IGPPVETIQRTTAR-DPIY-----PPDPPIKLKIQL	14870
QY	7348	FREPT---WPVGELGHAGLEITEESDDVALLAEAAGVKRKWKSPSRSLFHFPGRH	7403
Db	14871	ITYKNTVHLWSKKPPKNDG--GSPVTHYI--VECLAWDPTGTKKEAW-----	14911
QY	7404	LPLDEPAELGLSERVKASVEHISRILKGRPEGLEKEGPPRPKPGLASFRLSGLK---	SW 7459
Db	14912	-----RCNKRDVEEIQTVEDLVGEGEYE-----PRVAVNAAGVSK	14949
QY	7460	DRAPTFLRELSET-----VLQGSVTLACQVSAQAQAATWSKDGAPLSSSS	7507
Db	14950	PSATGVPTVVKDQCPSIDLKEFMEVEEGTDVNIVAKIKGVFPFTLTWFKAAPPKPDNK	15009
QY	7508	RVLISATLKNFOLL--TILVVAAE---DLGYVTCVSNALGTVTITGVLAKABRRPSSP	7561
Db	15010	EPLVYDTHVNLKLVDDTCTLVIPOSRSDTGLYITAVNNLGTSKEMRLNVLRGP---	15066
QY	7562	CPDIG-----EVYADGVLLVWKPVBESYGP---VTYIVQ-CSLEGGSWTTLASDFDCCYL	7612
Db	15067	-PPVGIPKFESVSADQMTLSWFPFKDGGSKITNVVIEKREANRKTWVHVSSPEKECTYT	15125
QY	7613	TSKLSRGGYTTRTACVSKAGMG-PYSSPSE	7642
Db	15126	IPKLEHGHEYVRIMAQNKYGIGEPLDSEPE	15156
 RESULT 15			
AAU05396			
ID	AAU05396 standard; protein; 26926 AA.		
XX	AAU05396;		
XX	24-OCT-2001 (first entry)		
DT	Human titin (connectin) protein sequence.		
XX	Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;		
KW	titin-related disease; zebrafish; heart failure; heart disease.		
XX	Homo sapiens.		
OS	WO200151666-A1.		
XX	19-JUL-2001.		
PN	12-JAN-2001; 2001WO-US001212.		
XX	12-JAN-2000; 2000US-0175787P.		
XX	(GENO ) GEN HOSPITAL CORP.		
PR	Fishman MC;		
XX			
PA			
XX			
PI			
XX			

QY 330 EPAVP-----FKKLODLEVREKESATFLCE---VP 357  
Db 1625 EIAAPELEPLHURYQOEQWEGDLYDKKQKPPFKKLTSLRUKRFGPAHFECRLTPIS 1684  
QY 358 OPSTEAAWFKBETRL-----WASAKYGI-----EEGTERRLT 390  
Db 1685 DPTMVVWLHDGKPLEAANRLRMINFEGYCSLDYGVAYSRDGSIITCRATNKYGTDTSA 1744  
QY 391 VRNYSADDDAYVYCETPEGR-----TVAEALVQGNL-----LEKLPRKT 430  
Db 1745 TLIVKDEKSLVEESOLPEGRGLORI BELERMAHEGALTGVTTDQKQKQPDIVLYPEPV 1804  
QY 431 AVRVDGTAMF-CVELAVPVGPHVHRLNOEEVAVGRVAISAEGRPHRLTISQCCLEDVQ 489  
Db 1805 RVLEGETARFCRVGTGYPOPKNVYLNQGLIRKSRFRVYDGI-HYLDIVDCSKSYDTGE 1863  
QY 490 VAFVAG-----DCOTSTRFCVSAAPRKPLQP-----PVD 518  
Db 1864 VKVTAENPEGVIEHKVKLEIQOQREDFRSLARAPEPRPEFHVHPEGKLOFEVQKVDREVD 1923  
QY 519 -----PVVKARWESSVILLSWPPPHGE-----RPVTIDGY-----LVEKKKLGITYTW 560  
Db 1924 TTEKEVVKLKAERI THEKVPESBELRSFKRTEEGYEAITAVELSKRKDSYBE 1983  
QY 561 I-----RCHAEBWATPELTVAD-----VABEGNFQFRVSALNSFGQSPYLEPFGTVHLAPKL 613  
Db 1984 LLRKTCDLHLW--TKELTEBEKKALAEKGITITPTFKPKIELSPSME-----APKI 2034  
QY 614 AVRPLKAVQAVEGEVTFSDLVASAG-----EMFLDQALKASS-VYEHICORTRH 666  
Db 2035 FERQSQVIG--QSDAHFRVRV-----VGKPDPECEWKYKGVKIERSDRIYTWPDENVC 2088  
QY 667 TLIREVPASLHGAKLXF-----VANGIESSIRMEVRAAGULTANKPPAAAAAREVLARLHE 722  
Db 2089 ELVIRDVTAE--DSASIMVKAINIAGETSSHAFLLVQAKLITFTQ-----ELQDVVAKEKD 2143  
QY 723 E-AQLLAELSQAANAATWLDGRITLSPGPKYEVQASAGRVLVVRDVARDDAGLYECV-- 779  
Db 2144 TMAFTECBTSPFVKVYKMDGMEVHEGDKYRMSDKVHFLSILITIDTSDAEDYSCVLV 2203  
QY 780 -SRGRIAYQLSVQGLARFLHKDMAGSCVDVAGGPAQFECETSEAHVHVHMYKDGMELG 838  
Db 2204 EDENVKTTAKLIVESAVVEFVKELQD--IEVPESYSGELECI VSPENIEGKWHNDVELK 2261  
QY 839 HSGERFLOEDVGTTHRLVAATVTRQDEGTYSRCVGEDSVDFRLRVSEPKVVFVFAKEOLAR 898  
Db 2262 SNGKYTTISRGR-RQNLTVKDVTKEDQGEYSFVIDGKTKTKLKM-KPRPIATLOGLSDQ 2319  
QY 899 KLQAEAGASATLSCEVAQAQTEVTVYKDGKLSKSSKVCMEATGCTTRLVVOQAGQADAG 958  
Db 2320 KVC--EGDIVOLEVKVSVLESVEGVMKDGQEVQPSDRHVHVDKQSHMLLIEDMYKEDAG 2377  
QY 959 EYS-----CEAGQRLSHLDVKEPKVFAKDQVAHSEVQAEAGANATLSCEVAQAQ- 1010  
Db 2378 NYSFTIPALGLSTGRVSVSDVITPL-----KDVNVIEGTKAVLECKVSPDV 2427  
QY 1011 AEVVMYKDGKLSLKVHVEAKCRRLRVVOQAGTKDAGDYSCAARGQVSRPLHITEP 1070  
Db 2428 TSVKXVYNDEQIKPDRVQALVKTGRKORLVINRTHASDEGPYKLI VGRVETNCLNSVEKI 2487  
QY 1071 KMMFAKESQVHNEVQAEAGASATLSCEVAQAQTEVTVYKDGKLSKSSKVCMEVKGCTTR 1130  
Db 2488 KIIRGLRDLTCTEQ-----NVVPEVELSHSGIDVLMFNKDXKIKPSSKYKIEBAHKIYK 2542  
QY 1131 LVLPQAKADAGEYSCAAGQVSRFHLHITPEKGVFAKESQVHNEVQAEAGITAMLSCEV 1190  
Db 2543 LTVLNMKDDGKTYTFVAGENMTSGKL--TVAGATISK--PLTDQTVAES-QBAVFECEV 2597  
QY 1191 AQPQTEVTVYKDGKLSKSSKVRMEVKGCTTRLVVOQAGADAGEYSCAAGQVSRFQHL 1250  
Db 2598 ANPDSKGEWLDRDGKHLPLTNIRSESDGKRRRIIAATKLDLDDIGEVTVKATSKTSAKUK 2657

QY 1251 ITEPKAVFAKEOLVHNEVTRTEAGASATLSCEVAQAQTEB-VTVYKDGKLSKSSKVRIBAA 1309  
Db 2658 V--FAVKIKKTL-KNLTIVTET-QDAVFTVETLHENVKGQWIKNGVVLENEKVAISVK 2712  
QY 1310 GCMROLVVOQAGQADAGEYTCIAG--GORLSFHLN-----VSEPKAVFAKEOLAHKVKQAE 1363  
Db 2713 GTIYSLRIKNCALVDES VYFRLGRLGASARLHVETVKI IKPKQDVTALEN----- 2763  
QY 1364 AGAIAATLSCEVAQAQ--TEVTVYKDGKLSKSSKVRMEAVGCTTRLVVOQACQADTGEYS 1421  
Db 2764 ----ATVAFESVSHDTPVVKMFHKSVEIKPSDKHRLVSEKRVKHLKMLQNTISPSDAGEYT 2819  
QY 1422 CEAGQRLSFLSDVAEPKVPVFAKEQFVH-----REVQQAAGASATLSCEVAQAQTEVVMY 1476  
Db 2820 AVVGQLECKAKLFV-----ETLHITKMKNIETPKTASFECEVSHFNVPSMWL 2869  
QY 1477 KDGKLSKSSKVRMEAVGCTTRLVVOQAGQADAGEYSCAAGQRLSFLHVAEPKAVPAK 1536  
Db 2870 KNGVSIEMSEKFKIYVQOKLHQLIIMTSTEDSABYTFVCGNDQVSATLTV-PIMITS- 2927  
QY 1537 EQPASREVOAEAGTATLSCEVAQAQTEVTVYKDGKLSKSSKVRMEAVGCTTRLVVOEA 1596  
Db 2928 ---MLKDINAEBEKDITFEVTVNYEGISYKMLKNGVEIKSTDKQCMRTKLTLSLNIENV 2984  
QY 1597 GOADAGEYSCAGDORLSFLHVAEPKVPVFAKEOPAHREVOAEAGASATLSCEVAQAQTE 1656  
Db 2985 HFGDAADYTFVAGKATSTATLYVEARHIEFRKH---INKDIKVLKRAMFECEVSEPDIT 3041  
QY 1657 VTVYKDGKLSKSSKVRMEAVGCTTRLVVOQAGQADAGEYSCAAGQRLSFLHVAEILEP 1716  
Db 3042 VOMKDDDELQITDRIKIQKEKYVHRLIIPSTRSDAGKTYVAGGVNSTAKLVEGEDV 3101  
QY 1717 QISERPCREPLVKEHEDIILTATLATPSAATVTLKDGVEI--RRSKRHETASQGDTH 1774  
Db 3102 RI--RSIRKEVQVIEKQRAVV--BFEVNEDDVDAAHWYKDGIEINFQVERHKYVVERIH 3157  
QY 1775 TLTVHGAQVLSAIVSCEVGAEGQDPVQVVE-EVAAKFCRLLEPVCGLGCTVTLACELS 1833  
Db 3158 RMFISETRQSDAGEYTFVAGRNRSVTLVYNAPPPQVLQELQVTVQSG-----K 3208  
QY 1834 PA--CAEYVWRCGNTQPRV-----GKRTQMAEGEVRSLTVILGFAEDAGEYVC 1880  
Db 3209 PARFCAMI--SGRPQPKI SWYKEEQLLSTGFKCKFLHDGQEYTLILLIEAFPEDAAVTC 3265  
QY 1881 ESRDDEH-----TSAQLTVSVPRVVK-----FMSGLSTVVAEGBEGEATFCQVVS 1925  
Db 3266 EAKNDYGVATTSSASLSVEVPEVSPDQMPYPPAIITPLOTDTVTSBGPAPFCRVSGT 3325  
QY 1926 DVAVVWFRDGLLOPSEKFAISQSGASHLTSIDLVLEDAGQITVEAGA-----SSSAAL 1981  
Db 3326 DLKVSWSYKDKKIKPSRFRMTQPEDITYLEIABAYPEDEGTYTFVANNAVGVQVSSITANL 3385  
QY 1982 RVREAPVLFKKLLBPQ-----TVBERSSVTVLELVR----- 2013  
Db 3386 SLEAPESILHERIEQEIEMENKEFSSSFLSAEBSGLHSAELQLSKINETLELLESVPYP 3445  
QY 2014 -----PWPELRWTRNATALAPKKNVEI 2035  
Db 3446 TKFDSKEGTGPIFKI VESNADISMGDVATLSVTVIGIPKXIQWFPNGVLLTTSADYKF 3505  
QY 2036 HAEGARHRLVHNVGFDADRGFFGCTETPDD-----KT 2066  
Db 3506 VFDGDDHSLIILFTKLEDEGEYTCWASNDYKGTICSAYLKINSKGEHKGKDTETESAVAKS 3565  
QY 2067 QAKL-----TVEMROVRLVRGLQAV-EAREOGTATMEVQLSHADVDSGTEDGRLRFQ 2117  
Db 3566 LEKJGCPCPHFLKELKPIRCAQGLPALFEYTVVGBEPAVTV-----TWFKENKQL- 3615  
QY 2118 QGPTC-----HLAVRGP--MHTLTLSGLRPEDSGLMWFAEGV-----HTSARLVV- 2161  
Db 3616 ---CTSYYTIIHNPGSGTFFIVNDPQREDSGLYICKAENMLGESCACAAELLVLEDDTD 3671  
QY 2162 -----TELPVPSFSR-PL-----QDVVTTTEKEKVTLECELSRENVDVRV 2198

Db 3672 MTDTPTCKAKSTPEAPEDPOTPLKGPAPVADLSQEIATFVKOTILKAALITEENQOOLS 3731  
QY 2199 -----LKQVELRAGKTWAIQAQ-----ACRSLTIY-----2225  
Db 3732 EHIKANELSSQLPLGAQELQSIQDKLTPESTRFLCINGSIHFQPLKEPSENLOQI 3791  
QY 2226 ---RCEPADQGVYVCDADAQS-----2244  
Db 3792 VQSQTFSKEGILMPEEPETOAVLSDTEKIFPSAMSEIQINSLTVEPLKTLAEPENYP 3851  
QY 2245 SASVKVQGRTY-----TLIYRRVLAE-----2265  
Db 3852 QSSIPEPMHESYLTVAEVEVLSKEKTVSDTNREQVTLQKQEAQASALLISQSLAEGHVES 3911  
QY 2266 -----DAGEIQVA-----ENAESRAQLRVKE-----2287  
Db 3912 LQSPDVMISQVNYEPLVPSEHSCTEGGKILIESANPLENAQDSAVRIEBSKSLRFLAL 3971  
QY 2288 -----LPVTLVRPLRDKIAMEKHGV-----2308  
Db 3972 BEKVLLKEESDNNVMPDQIIBSKREPVAIKKQVEVQGRDILLSKESLLSGIPEEORLN 4031  
QY 2309 LECQVSRA-SAQV-----RW-----2322  
Db 4032 LKIQICRALQAASEQPLFSEMLRTEKEVEAVNITQEPHIMCMYLVTSAKSVTEE 4091  
QY 2323 -----FKGSSQ 2327  
Db 4092 VTIITIEDVPMANJUMELRDAICAIYEEIDILTAEQPRIQOQAKTSLQEMDSFGSQ 4151  
QY 2328 BLQPKFYELSDGLYRLKLIISDVHAEDEDTYTCDAGV-----2366  
Db 4152 KVEPITEVESKYLISITEEVSYNVQSRVY-LDAPVTKGVASAVVDSKQDESCLKPS 4210  
QY 2367 ---KTSQOFFVEBOSIIVIGLQDVTVNEPAPA-----WFECETSIISVR 2408  
Db 4211 EKEBSESSESGTEEVATVKIQEAGGLIKEGPMIHTPLDVTVBEGDVLHATTSITNAK 4270  
QY 2409 PPKLLGLTQLQAGNVGLEGTVHRLMLAR-----TCSTMGPVHFVTKGSR 2458  
Db 4271 EVNWTFENKLVPSDEKFKCLOQONTYTLVIBKNTEDHQGEVCEALNDS-----GKTAT 4325  
QY 2459 SARL-VVSDIPVVLTRPLEPTKRELQSVLSCDFRPAPKA-VQMYKDDTPLSPSEPKFM 2516  
Db 4326 SAKLTWVRAAPVIRKTEPLEVALGHLAKTCTEIQSAPNVRFQWFRAGRIEYSDKCSI 4385  
QY 2517 SLEGQVABLRILRLMPADAGVYRCOAGSAHSSTEVTVZAREVTVTG-----BLQDA 2567  
Db 4386 RSSKYISSLLEILTQVDCGEYTCASNEYGSVSCIA-TLTVTPGGEKKVRLPKPKP 4444  
QY 2568 EATEEGWASFSCELSHEDDEEVWSLNGMPLYNDSFHEISHKGRRHITLVKSIQADAGIV 2627  
Db 4445 EPKEB--VVLKSVLRKRPPEEBEPKPKL-----EKVKKPAVPEPPPPKPVBEVPTV 4497  
QY 2628 RASSLVKSTSARLEVRVPVFLKALDLSAEERTLALQCEVSDPE-----AHVY 2678  
Db 4498 TKRERKIPETKVP-EIKPAIPALPAPEPKPPEAEVKTIPPPPEPEPTPIAAPVTPVW 4556  
QY 2679 WRKDGVLQPSDKYDFLHTAG-TRGLVVDVSPEDAGLYTCHVGS-----EETRARVRV 2731  
Db 4557 GKAAK---APKEEAAPKPIKGVPKKTPSPIAEERRKLPGSGGKEPPDEAPFTYQL 4613  
QY 2732 HDLHVGIITRLKTMVLE-----GESCSFECVLSHESADPAMWTVGKTVGSSSRFQATR 2787  
Db 4614 KAVPLKFVKIIDIILITSEFVGSSAIFELVSPSTAI--TTWMDKGSNIRESPKHFIA 4671  
QY 2788 QGRKYLIVREAPADGAGVSVR-----GLTSKASLIVREPAALIKPLEDO-WVAPGE 2842  
Db 4672 DGKDRKLHIIDVQLSDAGEYTCVLRGNKEKTSIAKLVVEELVRFVKTILEEVTVVKGQ 4731  
QY 2843 DYELRCELSRAGTPVHMLKDRK-AIRKSKYDVVCEGTWAMLVIRGASLKDAGEYTCVE 2901

Db 4732 PLYLSCELNKE-RDVVRKDGKIVVEKPGRIVPGVIGLMBALTIINDADDTDAGTYTVTVE 4790  
QY 2902 ASK-----STASHVEKANKCFTEELTNLQVBEKGTAVFTCKTEHPAATVTRKGLLELRA 2957  
Db 4791 NANNLESCSCVKVVEVIRDMLVKPIRDQHVKPKGTAFACDIKADTPNFKWFGYDEIPA 4850  
QY 2958 --SGKHQPSQEGTLRLTLISALEKADSDTYTCDIGQAQSAQLLVQGRVHIIEDELDVD 3015  
Db 4851 EPNDKTEIIRDNHLYLKNAMPEDIAEYAVEIEGKYPKAKLTIGEREVELLKPIEDYT 4910  
QY 3016 VQEGSSATFCRISIPANYEPVHFLDKTPLHANEINEIDAQPGGYHVLTLRLQALKDSOT 3075  
Db 4911 IYKESASDAEISEADI--PGOWKLKELLRSPTEIKAE--GGRFLLTLHVKLQDAGE 4968  
QY 3076 IYFEGDQDASALRVTEKPSVFSRELTDATITEGEDTLVCESTCDIPMCWTXDKGKTL 3135  
Db 4969 VLYQALNAITALLTVKEITELDFVPLKOVTVPERRQARPEC-VLTREANVWSKGPDI 5027  
QY 3136 RGSARCOLSHEGHRAQLITGATLQDSGRYKCEAGGACSSIVRVHARVRFQEAUKDLE 3195  
Db 5028 KSSDKFDIADGKKHILVINDSQFDDDEGVYTAEBEGKTSARLFTVIGIRLKEMSPLEDQT 5087  
QY 3196 VLEGGAATLRCLVSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLELVVRNLRPQDSGRYS 3255  
Db 5088 VKEGETATTVCELSHEKHVWVWFKNDAKLHTSRTVLISSEGGTKHLEMEVTLDDLSQIK 5147  
QY 3256 CSFGDOTTGATLTVTALPAQFICKLRNKEATEGATATLRCELSKTAPVWRKSETLRDG 3315  
Db 5148 AQVKELSSAQLKYLEADPYFTVKLHDKTAVEKDEITLACEVSKOVVWKFMDGEBIVPS 5207  
QY 3316 DRYCLRDQAMCELOIRGLAMVDAABYSCVCGEERTSASLTIRPMPAHFIRLHROESIE 3375  
Db 5208 PKYSIKADGLRILKIKKADLKDGKGYVDCGTDKTKANVTVEARLIEVEKPLYGVEFV 5267  
QY 3376 GATATLRLCELS-----KAAPVWRKGRSLRDGDRHSRLRQDQGVCELIQCLAVA 3425  
Db 5268 GETAHEIEISEPDPVGGWKLKQPLTASPDCEIIEGKX-----LILHNCQLGM- 5319  
QY 3426 DAGEYSCVCGEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLWCELSKAPV-EWRKG 3484  
Db 5320 -TGEVSFOANAKSAANLKVKEPLFITPLSDVKVFEKDEAKFECEVSREPRTFRLKG 5378  
QY 3485 PENLRDGRDRIYLBQETRCCELQICGLAMADAGEYLCVCGEERTSATLTIRALPARTEDV 3544  
Db 5379 TQBITGDDRPFLTKDGTGKSWIKSAFDEAKYFPEABDKHTSGKLIIEGIRLKEFTL 5438  
QY 3545 KQCEAEGEATVLOCEL-NSAAPVWRKGSSETLRDGRYSLRQDGTGKCELOIRGLANADT 3603  
Db 5439 KDVTAKEKSAVFTVELSHDNIRVWFKNDQRLHTTRSVSMQDEGKTHSITFKOLSIDDT 5498  
QY 3604 GEYSVCVCGEERTSAMLTVALPIKFTTEGLRNEEAETEGATAVLRCELSKM-APVEWKGHE 3662  
Db 5499 SQIRVEAMGMSSEPAKLTVEGDPYFTGKLQDVTGVEKDEVILQCEISKADAPVKWFKDGK 5558  
QY 3663 TLRDGRHSRLRQDAGCELOIRGLVAEDAGEYLCMGKERTSAMLTVRAMPSXFIIEGLAN 3722  
Db 5559 EIKPSKNAVIKTDGKRMILLKALKSDIGQYTCDCGTDKTSKGLDIEDREIKLVPLHS 5618  
QY 3723 EE-----3724  
Db 5619 VEVMETETARFETEISEDDIHANWKLKGEALLQTPDCEIKEBGKIHSLVHNCRLDQGTGG 5678  
Db 3725 -----ATEGDTATLWCELS--KAAPVWRKKGHEITL 3752  
Db 5679 VDFQAAVWKSASHLRVKPRVIGLLRPLKDVTTAGTATFDCBSLVEDIPVEWYLGKGL 5738  
QY 3753 RGDGRHSRLRQDGRCELQIRGLAVVDAGEYSCVCGEERTSATLTVALPARFIEDVKNQE 3812  
Db 5739 EPDKVVPVREGKVHTLTLRDVKLEDAAGEVQLTAKQFKTHANLFVKPEPPVEFKPLEDQT 5798  
QY 3813 ARGGATAVLOCELSKA-APVEWKGSETLRGGDRYSLRQDGTGKCELOIRGLSVADTGEYS 3871  
Db 5799 VEEGATAVLECEVSRENKVKWFKNGTTEILSKKXKIVADGRVVRKLVHIDCTPEDIKTYT 5858

QY 3872 CVCQERTSATLTYRAPQVREPRELOSIOABEGSTATLOCELSBPTATVWVSKGLOLQA 3931  
DQ 5859 CDADKFTSCNUNVVPVPHVFLRPLTDLOVREKEMARFECSELSRENKVMFKDGAIEKK 5918  
QY 3932 NGRREPRILQGGTAELVLQDLOREDTGEYCTTCGSOATSATITVTAAPVRFLRELQHOEVD 3991  
DQ 5919 GKXYDIIISKGAVRIILVINKCLLDDEAEYSCVETARTISGMLTVLEEEAVFTKLANIEVS 5978  
QY 3992 EGGTAHLCCELSRAGASVWFKGSLQLPCKAKYQWQDGAABELLVGVSOEDAGDYTC 4051  
DQ 5979 ETDITKLVCYSKFGAEVWTKGDEEIIETGRYILTEGRKRIILVIOAHLEDAAGNYNCR 6038  
QY 4052 TGHOTSMASLSVRPRPFKRLSLOEQETGDIARLCCOLSDAESGAVVOWLKEGVELHA 4111  
DQ 6039 LPSSRTDGRVKKVHELAAEFISKPNLEILEGEKAEFVCSIS--KESFPVQWKKDDKTLES 6096  
QY 4112 GPKYEMRSOGATRELLIHOLEAKDTGEYACVTGQKTAASLRVTEPEVTVIRGLVDAEVT 4171  
DQ 6097 GDKYDVIADGKRVLVVKVATLQDMGTVMVMGAAARAAHLVTE-KLRIVVPLKOTRVK 6155  
QY 4172 ADEDVPECSVSRAGATGVQVCLQGLPLQSNVTEVAVRDRGRIHTLRLKGVTPEDAGTVS 4231  
DQ 6156 EQEVEVFNCEVNTGAKAKWRNEATFDSSKY--IILQDOLVTLRIADHJDDQANYN 6213  
QY 4232 FHLGNH-----ASSAQLTVRAPEVITILPQDVQLSEGQDASFQCLRSRAGQEARWALG 4286  
DQ 6214 VSLNHRGENKSAANLIVEEDLRIEPLKDIETMEKKSVPFWCKVNRNLN-VTLKWTN 6272  
QY 4287 GVPLOANEMNDITVEQGLTHLITLHKVTELDAGTVSFHVTGCSSEAOLK-VTAKNVTVRG 4345  
DQ 6273 GEEVFFDNVRVYVDK-YKXMLTIDKCGFPDEGEYIVTAGODKSVAEILLIETAPTEVEH 6331  
QY 4346 LENVEALEGGALFECOLSQSEVAHTWLLDDEPVRTSENAEVVFFENGRLHLLLNKLNR 4405  
DQ 6332 LEDQTVTEFDDAVFSCQLSR-EKANVKWYRNGREIK--EGKYYKEFGDSIHLRIKDCR 6388  
QY 4406 PQDCSRVTFLAGDMVTSAFLTVRQWRLEILPLKNAAVRAGAQAARFTCTISEAVPVGEAS 4465  
DQ 6389 LDDECEYACGVEDSKSRARLFVEIPIIRPPODILEAFGADVFLAEILNK--DKVEVQ 6446  
QY 4466 WYINGAAVQDDSDWTVTADGSHQALLRSQAOPHAGEVTFACRDVASAR----- 4516  
DQ 6447 WLRNMWVVOGDKH-QMSEKGIHRLQICDIPKPDQGEYRPIAKDKARAKLELAAAPKI 6505  
QY 4517 ----- 4516  
DQ 6506 KTADQDLVVDVGKPLTMVVPYDAYPKAABWFKENEPLSTKTIDTTAEQTSFRILEAKKG 6565  
QY 4517 -----LTVLGLPDPPEDAEVAHSSHTVTLSWAAPMSDGGGLC 4555  
DQ 6566 DKGYKIVLQKHGKAGFNLKVIDVPGVRNLEVTETFDGEVSLAWBEEPLTDGGSKII 6625  
QY 4556 GYRVEVKEGATQWRNLCHLVPCEVVDGLAPGET-YRPRVAAGPVGAGEVPHLPQTV 4614  
DQ 6626 GYVVERDIIKRTWLTATDRAESCEFTVTGLQGVBYLFRVSARNRVGTGPEVETDNV 6685  
QY 4615 RLAEPPKVPVPP-----SAPESQVAGEDVSLELEVVAEAGEVIVHKG 4660  
DQ 6686 E-ARSKYDVOPPLNVTITDNRFGVSLTWEPPEYDGGAEITNYVIELRDKTSIRWDTAM 6744  
QY 4661 ERIOPGRFVVSQGRQMLVIKFTAEODQGEYHCGLAQSGSICPAATFOVALSPASVDE 4720  
DQ 6745 T-----VRAEDLSATVDVVEGGEYSFRVQNRIGVKG--PSAATPVKVADPIERP 6795  
QY 4721 APOSLPPEAAQEGDLHLHLEALARKMRSEBTLDSISELPEEDGSRQLPOEAEVAP 4780  
DQ 6796 SPPNLTSSDQTSQVQLKWEPPKLD---GGSTILGYIERCEE-GKDNWIRCNM-KLVP 6850  
QY 4781 DLSEGYSTABELARTGDADLSHTSSDDESAGTPSLVTVLTKKAGRPGTSLASKVGAPAA 4840  
DQ 6851 ELTVKVTGLEK---GNKLYRVSANENKAGVSDPSI----- 6883

QY 4841 PSVKPQQOQEPPLAAVRPPLGDLSTYD-LGDSMDXAAVK-----IQAAFKGYK 4887  
DQ 6884 -----LGPLTADDAFVEETMDLSAFKOGLEVIVPNPITILVPESTGY- 6924  
QY 4888 VRKEMKQOQEGPMFSHTFGDTEAQVGDALRLCECVASKADVARWLKDGVELTDCGRHHID 4947  
DQ 6925 -----PRPATWCFDGVLETGDRVKMKTLSA----- 6951  
QY 4948 QLGDGTCSLLIAGLORADAGCYTCVSN-----KFGQVTHSA 4984  
DQ 6952 -----YAEVLVIPSERSDKGYTLLENRVKTSIGEDVNVIAFPSAPKELKFDITKDS 7006  
QY 4985 CVVWGSSEABSSGGELDDAFRAARLRHLRFTK----- 5021  
DQ 7007 VHLT-----WEPDDDDGGFPLGYVVEKEVSRKWTVMDFVTDLFTVPDLVQCKEYLF 7062  
QY 5022 -----SPAEVSEELFSLADEGPAEPPEPADWQTYREDEHFI----- 5058  
DQ 7063 KVCARENKCGPEPAYV-DEPVNMSTPATVPOPPENVKWRDRTANSIFLTWDPKPKNDGSR 7121  
QY 5059 -----CIR-----FEALTEARQAVTRFQF--MFATLIGIGVEIKLVEQGPVRVEM 5100  
DQ 7122 IKGYIVERCPGSDKKWAGGEPVATKMEVITGLEBKWYA-----YRVT 7166  
QY 5101 CISKETRAPVVPPEPLPSLLTSDAAPVELTE--LQNEVQDQGYVSPFVTVGTGOMPMSVR 5158  
DQ 7167 LNRQASKPSRPTBIEIQAQVDTQEAPEIFLDVKLLAGLTVKAGTKIELPATVTKPEPKIT 7226  
QY 5159 WFKDCKLLEEDHYMINEDQGGHQLIITAVVPADMGVYRCLAENSMGVSTKLELVDL 5218  
DQ 7227 WTKAMILKQDKRITI-ENVPKKSTVTIVDSKRSOTGTIIEAVNVCGRATAVAVENVV- 7284  
QY 5219 TSTDYDTADATSESSYSAQYLLSSREQESTTDEGQLPVVEELRLDQVAPGTSLA 5278  
DQ 7285 DKPGPPAAFDITVNESCLLTWNPDRDGS----- 7316  
QY 5279 KPOLKVKYGPAPRLYFWKDGQPLTASAHIRMTGKKILHTLEIISVTRDSDQYAAYSNA 5338  
DQ 7317 ----KITNYVVER-----RATDSEVWHKL---SSTVKDTNFKATLIPN 7353  
QY 5339 MGAAYSSARLLVRGDEPEEKPSADVHQBOLVPRMLERFTPKYKKGKGS-SITFSVKVE-- 5395  
DQ 7354 KEYIFRVAENMYGAGEVQASPIAKYQDFPPGPTRELEPSDITKDAVTLTWCEPDDG 7413  
QY 5396 GRPVPTVHMLEEABRGVLWITGPTPGYTVASSAQHSLVLLDVGRHQGTY--TCIASN 5453  
DQ 7414 GSPI--TGYWV---ER-----LDPDTDKWVRCKMPVKDITTVRVKGLTNKKYRFRVLAEN 7464  
QY 5454 AAGQALCSASLHVSGLP-KVEEQEKVKEALISTFLQG--TTQALSAQGLETASFADLGGQ 5510  
DQ 7465 LAG-----PKPSKSTEPILIKDIDPPPPGKPTVKDVGKTSVR-----LNWT 7508  
QY 5511 RKEEPLAAKEALHLSLAEBVGTBEFLQKTSQITEMVSAKITQAKLOVPG-GDSDEDSKT 5569  
DQ 7509 KPEHDGGAIESYVTEMLKTGTDEWR-----VAEGVPTTQHLPLGLMEQGEYSFR 7559  
QY 5570 PSASPRHGRSPSSSIOBSSSESEDGDARGEIPDIYVVTADYLPGLAEQDAITLREGQY- 5628  
DQ 7560 VRANVKAESPE-----PSPVLCREKLYP 7586  
QY 5629 -----VEVLDAAH---PLRWLVTRKP-----TKSSPSRQGW--VSPAYLDRRLK 5667  
DQ 7587 PSPRWLEVINIKNTADLKWTVEKDGSGSPITNIVEKRDVRKQWQTVDTTVKDTKCT 7646  
QY 5668 LSP--EWG-----AARAPEPPGAVSEDEYKARLSSVIGELSSQAFVEELOFLOSHH 5719  
DQ 7647 VTPLTEGSXYFRVAEE-----NAIGQSDY-----TEIEDSVLAKDFTT----- 7685  
QY 5720 LQHLERCPHVPIAVAGKAVIFRNVDRIGRHFSSFLOEQCQDITDDVAMCFI-KNQAAP 5778  
DQ 7686 -----TPGPPYALA-----VVDVTKRHVDLKWEPKNDGGRPIQRYVIEKBERLG 7730  
QY 5779 EOYLEFLVGRV-----QAESVTVSTAIQEFYKXYAE-----EALLAGDPQ 5819

Db	7731	TRWK--AGKTAGDPCDPRVTDVTEGTEVQ--FOVRAENAGVCHPSEPTEIISIEDPTS	7786	Db	8744	-----HTWROPIETERS-KCDITGLLEBQ	8766
Qy	5820	PPPPLOHYLE-----QVER-----VQYQALLKELIRNKA	5851	Qy	6576	HYLHSHGVL-----HLDIKSPNIIIMVHPAREDIKICDFGFAQNTIPAELOQFSQVGS	6626
Db	7787	PPSPDLHVTDAGRKHIAIAWKEPKNGGSPIIYHVEMCPVGTKEKMRVNSRPKDLK	7846	Db	8767	ET--KFRVIANKKFCGPPVEIGP--ILAVDP-----LGP	8797
Qy	5852	RNRQNCALLEQAYAVVSALPORAENKLHVS-----LMENYFG-----TLEALGEPIROQ	5900	Qy	6627	PFVSPETIIQQNPVSEASDIWAMGVISVLSLTCSSPFFAGESDRATL-LNVLEGRVSWSSP	6685
Db	7847	FKVSEGVVPDKXY-----VLRVRANAICVSEPSISENVAKODPCDKPTID-----LETH	7897	Db	8798	P--TSPE-----RLTYT-----ERQSTITLTDWKEPRNGGSP	8828
Qy	5901	HFIYWEG-----APGARMPKGNHRHV-----FLFRN-----HLVICP-RRDS	5938	Qy	6686	MAAHUSE-----DAKOFIKATLQAPQAPSPAAQCULSHPWFLKSPMAEBAHFINTKOLKFL	6741
Db	7898	DIIVIEGKLSIPVFRFRAVPPTVSWHKGOGKVKASDRLTKMNDHISAHLEVPKSVRADA	7957	Db	8829	IOGYIIIEKRHRDKPDF-----ERVNKLCTPTTFLVNLDEHOMYFRVK-----	8873
Qy	5939	RTDIVSVFRNMKLSLIDNDQVEG-----DORAFEV-----WQ-----	5973	Qy	6742	LAERSWQRSLMSYKSLVNMRSIPELIRGPPDPSPISGVARHLCDRTGGSSSSSSSDNEIA	6801
Db	7958	GIYTIIT--LENKGSATASINVKVIGLPGCKOIKASDITKSSCKLTWEPPEPDGPTIL	8015	Db	8874	-----AVNEIGESEPSLP-LNVV-----IQDDEVP	8897
Qy	5974	-----EREDSVRKYLLQARTAIKGSW-VKEIC-----GIQORLALPV-	6010	Qy	6802	PFARAK-----SLP--PSPVTHSLP-----LHPRGLRPSASL--PEBAE	6837
Db	8016	HVLERREAGRTIPIVMSGENKLSWTVKDLIPNGEYFRVKA VNKVGGGEYTELKNPI	8075	Db	8898	PTIKLSVRGDTIKVYKAGEPVHIFADVTGLSPKIEWSKNETVIEKPTDALQITKEEVS	8957
Qy	6011	-----WRPDP--FEEELADCTAELGETVKLACRVCTGTPKPVLSWYKDGKAVQVDPHHILIE	6064	Qy	6838	ASERSTEAPAPPASPEGAGPPAAQGCVPRHVSIRSLFVHQAGESPEHGALAPGSRHHPAR	6897
Db	8076	AQDPKQPPDPVDVEVHNPTAE-----AMTITWKPLL--YDGSXIM--GYIIEK	8121	Db	8958	RSEAKTELIPKAVREDKGTIVYTASNRLGVSFRNVHVEYDR-----PSPRENLA	9009
Qy	6065	DPDGS-----CALILDSLTGVDSG--QWCPFAASAAGNCSTLKGILVQVPP-RF	6110	Qy	6898	RHLLKGGYIAGALP--GLREPLMEHRVLEBAAREQA-----TLLAK-----A	6940
Db	8122	IAKEERKRCNEHLVPLITLTAKGLEGEKEYQFRVAENAG-----ISEPSRATPTTKA	8177	Db	9010	TDIKAESCVLTWDAPLDNGGSE--ITHYVIDRDASRKKAWEVEVNTAVEKRYGWKLI	9067
Qy	6111	VNKVRASPV-----EGEDAQFTCTIEGAPVQIRWYKD-----GALLTT	6150	Qy	6941	PSFETALRLPASGTHLAPGSHSHLEHDSPTPRPSEACGEAORLPSPAGGAPIDMGH	7000
Db	8178	VDPIDAPKVLRTSLVLEKRGDEIALDASISGSPYPTITWINDENVIVPEEIKKAAAPLVR	8237	Db	9068	PNQGYFRVRANVKY--GISDECKSDKVIQDP-----YRLPGPP--GKP-KVLAR	9113
Qy	6151	GNKFO-----TLSEPRCLLVIVRAASKEDGLYECELVNLKLSARASA	6195	Qy	7001	PQSGKQLPST-----GGHFETAQ--PERPSPDSF-MQOPAPFFCHPKQG-----SAPQ--E	7045
Db	8238	RRKGEVQEEBEPVLPULTORLSDINSKGESQLRVRDSLRDPHGLYMIKVENDHGIAKAPC	8297	Db	9114	TGKSMVSTPPLDNGGSPITGVWLEKREEGSPYMSRVSRAPIITKVGLKGVFNVRILLE	9173
Qy	6196	ELRIGS-----PMLQAOECHREQLVAAYE-----DTTLERADQE-----VT	6232	Qy	7046	GCS--PHPAVAPCPGFPFGSCKEAPLVFSSP-FLCQPOAPAPAKASPLDSEKMGPGDI	7103
Db	8298	TVSVLDTGPPINFFVEDIRKTSVLCKWEPPLDGSGSEIINYTLKDKTKPDSEWIVVT	8357	Db	9174	GVKYQFRAMAINAAGTGPPE-----PSDEVAGDPIFFPPGP-PSCBEVKDKT-KSSI	9224
Qy	6233	SVLK-----RLLOPK-----ARGPS-----TGDLTGCPCP-----	6258	Qy	7104	SLPGRPKPGPCSPGSPGASOASSQSLSRVGSSQVGTBPSPSLDABGM--TOEAEPLDS	7161
Db	8358	STLRCHKYSVTKLIEGKBYLFRVRAENRFGPPPCVSKPLVAKOFFGPPDAPDKPIVEDV	8417	Db	9225	SLGKWP--PAKGG-----SPIKGIYVMEQECT-----TDMKRVNEPDKLIT	9266
Qy	6259	-----RGAPAL-----	6264	Qy	7162	-----TPTLQPOQATWRKFSLGGRGYAGYGTFFAGDGAGMLGQGPWARIATAWAV	7217
Db	8418	TSNSMLVKNNEPKDNGSPILGVWLEKREVNTHSRVNSLLNALKANVDGLLEGLTVYF	8477	Db	9267	CECVWPNLKE-----LRKYRF-----RVK-AV	9287
Qy	6265	-----QBTGSOPTVTGSEAPAVPRVPQPLHLEGPEQPEAIARAQSWTVPIRMGAAPWG	6321	Qy	7218	SQSEEEQEEARAESOSEEQEAREABSPLOVSARVPPEVG-----RAPTRS	7264
Db	8478	RVCAENAAFP--GKFSPPSDPKTAHDPISEPGP--PIPRVTDTSSTTIELEWEPFAPNG	8532	Db	9288	NEAGESEPSITTGEIIPATDIOEB-----PEVFI-----DIGAQDCLVCKAGSQIRIPAVI	9337
Qy	6322	AG-----TQELWD-----VSHSVVRETTQ-RTYTYQAIDTHTA-----	6354	Qy	7265	SPBFTP--WEDIGQVSLVQ--IRDLSDGDAEADTISLDSIVDPAYNLNLSLYDI--K	7316
Db	8533	GGEIVGYFVDKQJLVGNTKWSRCTEKMKVQYTVKEIREGADYKLRVSANVAAGEGPPGE	8592	Db	9338	KGRPTPKSWEFPDGKAKAMKQGVHIDIPEDAQLETAENSSVIIIPBCKRSHTKYSITAK	9397
Qy	6355	-----RPPSMQVTIE--DVQAGTGTGAQFEAIEBGPDPQSVTVYKDSVOLVDST	6401	Qy	7317	YLPFSFWIFKVPKSAQPEPPSPMAEELAEFF-EPTWPKPEGLG-----	7360
Db	8593	TQPVTVAEQPEPAVELDVSVKGGIQAAGKTLRIPAVTVGRPVPTKVTKKEGEL-DKD	8651	Db	9398	NKAGQKTANCRVKMVDVPGPKDKLVKZSDITRSCRLSKWMPDDDDGDRKGYVIEKRTID	9457
Qy	6402	RLSQOQEGTYSVLVRHVASKADAGYVTCIAQNTGQVLCBAELLVLGDNBP-----DSEK	6457	Qy	7361	-----PHAGL-----EITESEEDVDALLAAEAAGRKRKWSPPSRSLFHPGRLPL	7406
Db	8652	RVIDNVGPKSLIITKDALRKDHGRVITATNSCGSKFAAARVEVF--DVPGPVLDDKP	8708	Db	9458	GKAWTKVNPDCGTTTFVVPDLLSEQQYFRVRAENRFGIGPPVETIQTARTDP--IYPP	9515
Qy	6458	QSHRRK--LHSPFYEKBEIGRGVFGVKRQVQKGNKILCAAKFIFLRSRTRAQAYREDI	6515	Qy	7407	DEPAELGLRERVKAASV-----PHISRLKGRPEGLEKEG-----	7440
Db	8709	VVTRNKNLLNWSDEDDGSGSITGFI--IERKDKM-----	8743	Db	9516	DPITKLKIGITKNTVHLSWKPKNDGSPVTHYVECLAWDPTGTKEAWRQCNKRDVE	9575
Qy	6516	LAALSHPLVTGLLDQFETPKTKLILILELCSSEELDLRLYRKGVVTEAEVKVIOQLVEGL	6575	Qy	7441	-----PPRKKPGLASFLSGLKSW-----DRAPTF-LRELSDETIVL	7476
				Db	9576	ELOQTVEDLVEGGEYFRVKA VNAAGVSKPSATVGCDCQPDMPSPSIDLKERFE--VBE	9633





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 11:13:38 ; Search time 148 Seconds  
(without alignments)  
5178.744 Million cell updates/sec

Title: US-10-077-130-5  
Perfect score: 41273  
Sequence: 1 MDQPSGAPFLTRKAFV.....RNREKRALLYKHNLAQVR 7968

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_78.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3577.5	8.7	26926	1	I38344
2	3313	8.0	7962	1	I38346
3	2629	6.4	4162	2	T42633
4	2489	6.0	6642	2	T29757
5	2048	5.0	5198	2	T32290
6	2043.5	5.0	5175	2	T20992
7	1636.5	4.0	7160	2	T27935
8	1634	4.0	6839	2	S57242
9	1619	3.9	6831	2	A88852
10	1392	3.4	6658	2	T13931
11	1074	2.6	4391	2	A38096
12	1046.5	2.5	6805	2	S20901
13	1012	2.5	1906	1	S68235
14	963.5	2.3	2783	2	T34416
15	869	2.1	1323	2	PN0568
16	786.5	1.9	3488	2	T34418
17	767	1.9	3707	2	S18252
18	703.5	1.7	1398	2	T25568
19	645.5	1.6	1176	2	JN0583
20	622	1.5	1147	2	A59307
21	571	1.4	1274	2	S50500
22	532	1.3	1694	2	S50065
23	528.5	1.3	1132	2	A35089
24	520	1.3	1142	2	A36845
25	519	1.3	3375	2	T19821
26	501.5	1.2	608	2	A35021
27	480.5	1.2	610	2	A28788
28	472	1.1	1896	2	T08851
29	465.5	1.1	2541	2	T29340

RESULT 1

I38344

titin, cardiac muscle [validated] - human

N;Alternate names: Connectin

N;Contains: serine/threonine-specific protein kinase (EC 2.7.1.-)

C;Species: Homo sapiens (man)

C;Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 15-Sep-2000

C;Accession: I38344; I38345; S20898; S20897; S20899; S63665; S37393

R;Label: S.; Kolmerer, B.

Science 270, 293-296, 1995

A;Title: Titins: Giant proteins in charge of muscle ultrastructure and elasticity.

A;Reference number: A57430; MUID:96026330; PMID:7569978

A;Accession: I38344

A;Status: nucleic acid sequence not shown; translation not shown; translated from GB/EM

A;Molecule type: mRNA

A;Residues: 1-26926 <LAB1>

A;Cross-references: EMBL:X90568; NID:G1017424; PID:G1017425

R;Musco, G.; Tziatzios, C.; Schuck, P.; Pastore, A.

Biochemistry 34, 553-561, 1995

A;Title: Dissecting titin into its structural motifs: Identification of an alpha-helix

A;Reference number: I38345; MUID:9519041; PMID:7819249

A;Accession: I38345

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1977-2014 <MUG>

A;Cross-references: EMBL:X83270; NID:G602579; PIDN:CAA58243.1; PID:G602580

A;Note: conformation and properties are reported for a synthetic peptide corresponding

R;Label: S.; Gauter, M.; Lakey, A.; Trinick, J.

EMBO J. 11, 1711-1716, 1992

A;Title: Towards a molecular understanding of titin.

A;Reference number: S20897; MUID:92258380; PMID:1582406

A;Accession: S20898

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 13597-14200, 'I', 14202-14696 <LAB2>

A;Cross-references: EMBL:X64698; NID:G37192; PIDN:CAA45939.1; PID:G37193

A;Accession: S20897

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: mRNA

A;Residues: 16330-16392, 'S', 16384-16756, 'F', 16758-16860 <LAB3>

A;Cross-references: EMBL:X64699; NID:G37190; PIDN:CAA45940.1; PID:G37191

A;Accession: S20899

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: mRNA

A;Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 22483-

A;Cross-references: EMBL:X64697; NID:G37190; PIDN:CAA45938.1; PID:G37195

R;Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labelt, S.

J. Mol. Biol. 256, 556-563, 1996

A;Title: Genomic organization of M line titin and its tissue-specific expression in two

A;Reference number: S63665; MUID:96177761; PMID:8604138

A;Accession: S63665

A;Status: nucleic acid sequence not shown

A: Molecule type: DNA  
 A: Residues: 26729-26825 <KOL>  
 A: Cross-references: EMBL: X92412; NID: g1236761  
 R: Gautel, M.; Leonard, K.; Labbett, S.  
 ENBO J. 12, 3827-3834, 1993  
 A: Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentiating muscle  
 A: Reference number: S37393; MUID: 94008990; PMID: 8404852  
 A: Accession: S37393  
 A: Molecule type: mRNA  
 A: Residues: 26831-26926 <GAU>  
 R: Improta, S.; Politou, A.S.; Pastore, A.  
 submitted to the Brookhaven Protein Data Bank, February 1996  
 A: Reference number: A66736; PDB: 1TIT  
 A: Contents: annotation; conformation by (1)H-NMR, residues 5253-5341  
 R: Pfuhl, M.; Pastore, A.  
 submitted to the Brookhaven Protein Data Bank, August 1996  
 A: Reference number: A66201; PDB: 1NCT  
 A: Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155  
 C: Genetics:  
 A: Gene: GDB: TTN  
 A: Cross-references: GDB: 127867; OMIM: 188840  
 A: Map position: 2q31-2q32  
 C: Function:  
 A: Description: structural protein forming filaments in striated muscle  
 C: Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pro  
 C: Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco  
 structural protein  
 F: 24752-25008/Domain: protein kinase homology <KIN>  
 F: 84, 177, 905, 2276, 2378, 2459, 2481, 2563, 2669, 2763, 2896, 3088, 3179, 3384, 3432, 3628, 3772, 4068,  
 98, 11066, 11488, 11515, 11635, 11949, 12170, 12478, 12526, 12645, 12875, 13001, 13036, 13295, 13540, 1  
 ratus predicted  
 F: 16780, 16976, 17579, 17602, 17667, 17681, 17845, 17899, 18121, 18188, 18209, 18336, 18670, 18680, 18  
 .21900, 21935, 22495, 22627, 22897, 23024, 23318, 23883, 24012, 24177, 24290, 24447, 24642, 248  
 F: 26171, 26178, 26184, 26190/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 8.7%; Score 3577.5; DB 1; Length 26926;  
 Best local similarity 20.2%; Pred. No. 5.2e-97;  
 Matches 1929; Conservative 1288; Mismatches 3662; Indels 2663; Gaps 320;

QY 4 PQFGAPRFLTRPKAFVVSVGKDATLSCQIVGNPTQVSWKQDQPPVTAGARFLA---Q 60  
 DB 1030 PGEPAAPYFIKPVVQKLVGGVGVGQGNPKPHVWKSGVPLTTGYRYKVSYNKQ 1089  
 QY 61 DGLYRLTILDALGDSQYVCRANAIGAFPAVGI-----QVD 100  
 DB 1090 TGEC-KLVISMTFADDAGETIIVRNKHGETSASALLEADYELLWKSOEMLYOTQVT 1148  
 QY 101 A---EACACAEQAPHL-----113  
 DB 1149 AFVQEPVGETAPGVFVSEYEKEQALIRKKMAKDVTVVRTYVEDQEFHSSPEERL 1208  
 QY 114 -----LRPTSIRVREGSATRCRV 133  
 DB 1209 IKEIYRIIKTLEELDEEKEKAVDISESAVESGPDRLRNKRYILEGMGVTTHCKM 1268  
 QY 134 GSPRPVAVSKQGRRLGEPDGPVRRVVEELGEASA-LRIRAAAPRQDGTVEVRAENPLGA 192  
 DB 1269 SGYPLPKIAWKDGKRI--KHGERYQNDFLQDGEASLIRIPVLPDEGIYTAFAINIKGN 1326  
 QY 193 ASAAALVWSDAADTASR--PGTSTAALLAHQRR---REANRAGAPASPP-----240  
 DB 1327 AICSGKLYVEPAALGAPTPIPTLEPVSRIRSLSPRSVSRSPIRMSPARMSPARMS 1386  
 QY 241 -----STGTR-----TCTVTEGKHARLSCVVTGEPKPEPTVWKK 273  
 DB 1387 SPARMSFGRLLEETDESQERLYKPFVFLKPVFKCLEGANCDFDLKVGVRPNPETFWEH 1446  
 QY 274 DQQLVTEGRRH--VYEDAGNFVILFKQSDRGLYTCTANLVGQTVSSVLVV-----328  
 DB 1447 DGOQIVNDYTHKVIKEDGTQSLI--IVPATPSDGEWTVVAQNRAGRSSISVILTVEAV 1504  
 QY 329 -----328

DB 1505 EHQQVPMFVEKLVKNVNIKEGSRLEMKVTRATGNPNPDIIVLKNSDIIVPHKYPKIRIEGTX 1564  
 QY 329 -----R 329  
 DB 1565 GSAALKIDSTVSDQSAWYATATANKAGRDTTRCKVNVVEFAEPPEPKLIIPRGTYRAK 1624  
 QY 330 EPAVF-----FKKRLQDLEVRKEKESAFPLCE---VP 357  
 DB 1625 EIAAPELSPHLRYQEQEOWEEDLYDKKQKQPFKKKLTSLRLKRGFPAHFECRLPIS 1684  
 QY 338 QPSTEAAWFKBETRL-----WASAKYGI-----EEGTERLRT 390  
 DB 1685 DPTWVVEMLHDGKPLEAANRLRMINEFGYCSLDYGVAYSRSDSGITCRATNKYGTDTSTA 1744  
 QY 391 VRNVSADDDAVVICPTPESSR-----TVAEIAYQGNL-----LRKLPRKT 430  
 DB 1745 TLIVKDEKSLVEESQLPGRKGLQRIEELERVAHEGALTGVTTQKQKQKQPDIVLYPEPV 1804  
 QY 431 AVRVDGTAMP-CVELAVPVGPVHMLRNOQEEVVAGRVVAISAEGRTHLTITISQCCELDVQ 489  
 DB 1805 RVLEGETARFCRVGTGYPQKVNWYLNQGLIRKSRFRVRYDGI-HVLDIVDCKSYDTGE 1863  
 QY 490 VAFMAG-----DCQTSRFCVSAPRKPLOP-----PVD 518  
 DB 1864 VKVTANPEGVIEHKVLEIQOEDRFSVLRRAPEPRPEFHVHPEPGKQFEVQKVDPRVD 1923  
 QY 519 -----PVVKARMESSVILSNPSPHGE-----RPVTIDG-----LVEKKKLGTYTW 560  
 DB 1924 TTETKEVVKLRABRIITHEKPESEELRSKFKRTEEGYEAITAVELSKRKKDESVEE 1983  
 QY 561 I-----RCEAEAWATPELTIVAD---VAEEGNQFQVRSALNSFGOSPYLEPGTIVHLAPKL 613  
 DB 1984 LIRKTKDELLHW--TKELTEEEKALABEGKITPTFKPKDIELSPME-----APKI 2034  
 QY 614 AVRTPLKAVQAVEGGEVTFSDLTAVASAG-----EMFLDQQAALKASS-VYEIHCDRTRH 666  
 DB 2035 PERIOSQIVG--QGSDAHFRVRV---VGKPDPECEWVKGVKIERSDRIYVWYVPEDNVC 2088  
 QY 667 TLITREVPASLHGAKLKF-----VANGIESIRMEVRAAPGLTANKPPAAAREVLARLHE 722  
 DB 2089 ELVIRDVTAB--DSASIMVKAINIAGETSSHAFLVQAKLITFTQ-----ELQDVVAKEKD 2143  
 QY 723 E-AQLLAELSQAQAAVTWLKDGRTLSLPGCKPEVOASAGRRVLLVRDVARDDAGLYECV-- 779  
 DB 2144 TWAFECTSPEFFVKVKKYKQGMVEHEDGKYMESDRKVFHLSLTITDTSADAEYSVLV 2203  
 QY 780 -SRGRIAYQLSVQGLARFLFKHMAGSCVDVAGGPAQFECETSEAHVHVHWKGMELG 838  
 DB 2204 EDENVKTTAKLIVGAVVFEVKELQD--IEVPESYSGELECIVSPENIEGKWYENDVELK 2261  
 QY 839 HSGERFLOEDYGTGHRLLVAATVTRQDSEGTSCRVGSDSDVDFLRVSRPKVVFVFAEQIARR 898  
 DB 2262 SNGKTYITTSRGG-KQLNTVTKDVTKEGQSEYFVTDGKTTCKLAW-KPRPIAILQGLSDQ 2319  
 QY 899 KLOAEAGASATLSCEVAQAQTEVTYKDGKLSLSSSKVCMETACTRRLVYVQQAQADAG 958  
 DB 2320 KVC--EGDIVQLEVKVLSLEVEGVMKDGQVQSPDRVHIVIDKQSHMLLIETDMTKEDAG 2377  
 QY 959 EYS-----CEAGQORLSFHLQVKEPKVVFPAKQVAHSEVQAEAGANATLSCEVAQAQ- 1010  
 DB 2378 NYSFTIPALGUSTSRGVSVSDVTITPL-----KQNVNIEGKVALECKVSDVDP 2427  
 QY 1011 AEVVMYKDGKLSLSSKLVHVEAKGRRRLVYVQQAQKTDAGDYSCEARGQVSRFLHITEP 1070  
 DB 2428 TSVKWLNDQIKPDDRVAQAVKGTQKRLVINRTHASDEGPKLVIGRVETNCNLSVEKI 2487  
 QY 1071 KMFVAKESGVINEVQAEAGASAMLSCEVAQAQTEVTYKDGKLSLSSSKVCMVEKGTTR 1130  
 DB 2488 KIIRGLRLDTCTETQ-----NVVFEVLSHSGIDVLNFKDKKEIKPSKYYKIEAHGKIYK 2542  
 QY 1131 LVLPQAGKADAGEYSCBAGGQVRSFHLHITPEKGVFAKEQSVHNEVQAEAGTAMLSCEV 1190  
 DB 2543 LTVLNMKDDSGKTYTFVAGENWTSGLK--TVAGGAISK--PLTDQTVAES-QEAVFECEV 2597





QY	4781	DLSEGYSTADELARTGDADLSHTSSDDESRAGTSLVLYLKAGRPOTSPILASKVGAPAA	4940
Db	6851	ELTYKVTGLEK---GNKYLRVSAENKAGVSDPSEI	6883
QY	4841	PSVKPQQOQEPALAAVRPLGLDLSKD-LGDSBMDKAAVK	4887
Db	6884	-----LGPLTADDAFVEFTMOLSAFKDGLVIVPNPITILVPSGTY- 6924	
QY	4888	VRKEMKQOEGPMFSHTFGDTBAQVGDALRLSCVVASKADVARWLKDGVELTDCGRHHIID	4947
Db	6925	-----PRETATWCFGDKVLETGDRVKMKTLSA----- 6951	
QY	4948	QLGDTGCSLLIAGLDRADAGCYTCQVSN-----KFGQVTHSA	4984
Db	6952	-----YAEVLISPSERSDKGIYLUKENRVTTISGEIDNVIAIRPSAPKELKFGDITKDS	7006
QY	4985	CVVYSGSSEASESSGGDLDAFRAARLRLFRFK----- 5021	
Db	7007	VHLT-----WEPPDDGGSLPGYVVEKREVSRTKTKVMDFTVLEFTVPLVQGEYLF	7062
QY	5022	-----SPAVSDEELFLSADGPAEPADWOTVREDEHFI----- 5058	
Db	7063	KVCARNKCGPEPAYV-DEPNMSTPATVPDPENVKWRDTANSIFLTWDPKNDGGSR	7121
QY	5059	-----CIR-----FEALTARQAVTRFOE--MFATLGIGVEIKLVEQGRPRVEM	5100
Db	7122	IKGYIVERCPGSKWACGPEVAETKMEVTGLEBKWYA-----YRKVT	7166
QY	5101	CISKETPAVVPPEPLPSLLTSDAARVFLTE--LQNEVQDQPVSVFDCVVGTGPMPSVR	5158
Db	7167	LNROGASKPSRPTBIEIQAQVTEAPEIFLDVKKLAGLTVKAGTIELPATVTVGPEPKIT	7226
QY	5159	WFKDGLLEEDHYMINEDQGGHOLIITAVVPADMGVYVCLAENMGVSSTKAEKLVDL	5218
Db	7227	WTKADMILUKQDKRITI-ENVPKSTVTVDSKRSDTGYIIEAVNVCGRATAVENVV-L	7284
QY	5219	TSTDYDTAADATESSEYSAQGYLSSREQESTTDEGOLPVQVVEELRDLQVAPGTRLA	5278
Db	7285	DKPGPAAAFDITVINESCLLTWNPDRDGGG----- 7316	
QY	5279	KFOLKVKGYPAPRLYFWKDGQPLTASAHIRTKKILHTLEIISVTRSDSQYAAYSNA	5338
Db	7317	-----KITNYVVER-----RATDSEVWHKL---SSTVKDNTNFKATKLIPN	7353
QY	5339	MGAAYSSARLLVRGDEPEEKPASDVHQLVPPRMLERFTPKVKVKGSS-SITFSVKVE--	5395
Db	7354	KEYIFRVAENMYGAGEPVQASPIAKYQFDPGPPTELEPSDITKOAIVTLTWCEPDDDG	7413
QY	5396	GRPVPTVHMLEEAREGVNLWIGPTDPGYTVVASSAQOQSHLVLLDVGRHQGTY--TCIASN	5453
Db	7414	GSPI-TGYVW-----ER-----LDPDTDKVRCKNMPVKDITYRVKGLTNKKYRFRVLAEN	7464
QY	5454	RAQALCASLHVSGLP-KVEBEQVKREALISTFLOG--TTQATISAQGLETASFADLGGQ	5510
Db	7465	LAG-----PKPKSKSTEPILIKDIPDPHPGPPPTVKDVKTSVR-----LNWT	7508
QY	5511	RKEEPLAAKEALHLSLAEVGTETFEFLQKLTQITBMWSAKITQAKLQVFG-GDSDEDSKT	5569
Db	7509	KPEHDGGAKIESYVIELMTKTGDWVR-----VAEGVPTTQHLPLPGLMGEGYSFR	7559
QY	5570	PSASPRHGESRPSISIQSSSESESDGARGIFDIYVVTADYLPFGABODAITREGOY- 5628	
Db	7560	VRVANKAGESEFSE-----PSDVPVLCREKLYP 7586	
QY	5629	-----VEVLDAAH---PLRMLVTRTKP-----TKSSPSRQGW--VSPAYLDRRLK	5667
Db	7587	PSPPRWLEVINITKNTADLKWTVPEKDGSPITNIVIEKRDVRRKGWQTVDTTKDTKCT	7646
QY	5669	LSP--EWG-----RAEAPPEPGCEAVSEDEYKARLUSVIOQLLSEQAQFVEELOFLOSHH	5719
Db	7647	VTPLTEGSLYVPRVAE-----NAIGSDY-----TEIEDSVLAKDTFT----- 7685	
QY	5720	LQHLCRCHPVPIAVAQKAVIFRNVRDIGRHFSSFLQELQQCQDQDDDDVAVCFI-KNOAAF	5778
Db	7686	-----TPGPPYALA-----VVDVTKRHDVLKWEPPKNDGGRPIQRYVIEKBERLG	7730
QY	5779	EQYLEFLVGRV-----QAESVYVYSTAQEYFKYAE-----EALLAGDPSQ	5819
Db	7731	TRWVK--ACKTAGPCDNFRVTDVIEGTEVQ--FQVRAENEAGVGHPSBTEILSIEDPTS	7786
QY	5820	PPPPPLQHYLE-----QOVER-----VQRYOALLKELIRNKA	5851
Db	7787	PPSPQLDLHTDAGRKHIAIAWKPEKNGSPPIGYHVEMCPVGTCKMWRVNSRPIKDLK	7846
QY	5852	ENRQNCALLBQAVVVSALPORAENKLHVS--LMENYFG-----TLEALGEPIROG	5900
Db	7847	FKVEEGVDPKEY-----VLRVRAVNAIGVSEBSEISENVVAXDPCKPID-----LETH	7897
QY	5901	HFTVWEG-----APGARMPWKGHNHIV-----ELFRN-----HLVICKP-RRDS	5938
Db	7898	DIIVIEGEKLSIPVPFRAVFPVTVSMHKDGKEVKASDRLTMKNDHISAHLEVPKSVRADA	7957
QY	5939	RTDTVSVVFRNMKLSIDLNDQVEG-----DDRAFEV-----WO----- 5973	
Db	7958	GIYTIIT--LENKLSATASINVKVIGLPGPCDKIDKASDITKSSCKLTWEPFEDGTPIL	8015
QY	5974	-----EREDSVRKYLLQARTAIKSSW-VKEIC-----GIOORLALPV- 6010	
Db	8016	HYVLEERAGRTYIPVMGGENKLSMTVKDLIPNGEYFPFRKAVNKVGGGEYIELKXPI	8075
QY	6011	-----WPPD--FEEELADCTAELGEIVKLAACRVGTGPKPVI-SWYKDGKAVQVDPHILIE	6064
Db	8076	AQPKPPPPPPVDEVHNPTAE-----AMTIWKPPL--YDGGSKIM--GIIIEK	8121
QY	6065	DPDGS-----CALLIDSITGVDSG--CYMCFAASAAGNCSTLGIILVQVPP-RF	6110
Db	8122	IAGGEERWKRCEHNLVPIITYTAKGLEEGKEYQFRVRAENAAG--ISEPSRATPPTKA	8177
QY	6111	VNKVRASPV-----EGEDAQFTCTIEGAPYPOIRWYKD-----GALLTT	6150
Db	8178	VDPIDAPKILRTSLEVRKRGDEIALDASISGSPYTTITWIKOENVIVPEIKKRAAPLVR	8237
QY	6151	GNKFO-----TLESPRSGLLVIRAAKEDLGLYECLELVNRLGSRASA	6195
Db	8238	RRKGEVQEBEPFVLPIQRLSIDNSKKGESQLRVRDLSRPHGLYMIKVENDHIGIAKPC	8297
QY	6196	ELRIQS-----PMLQAQEQCHREOLVAVE-----DTTLERADQE-----VT	6232
Db	8298	TVSVLDTGPPPIINFVEDIRKTSVLCKWBPPLDGGSEIINYTLKKDKTKPDSEWIVVT	8357
QY	6233	SVLK-----RLLGPK-----AGPS-----TGDLTGPGPCP----- 6258	
Db	8358	STLRHCKYSVKTLIEKEYLFRVRAENRRPGPPPCVSKPLVAKDPGPPDADPKPIVEDV	8417
QY	6259	-----RGAPAL----- 6264	
Db	8418	TSNSMLVKXNEPKDNGSPILGYWLEKREVNSTHWSRVNKSLLNALKANVDGLLEGLTYVF	8477
QY	6265	-----QETGSPPTVTGTEAPVAPVPRVQPOLHHEGPEQEPERAIARAQEWTVPIRMEGAWPG	6321
Db	8478	RVCAENAAAG--GKFSPPSDPKTAHDP:SPGPF--PIPRVDTTSSTTIELEWEPFAFNG	8532
QY	6322	AG-----TGBLLWD-----VHSHVVVRETTQ-RITYYQAIIDHTA----- 6354	
Db	8533	GGELVGYFVDKQLVGNKMSRCTEKMKVRYQYTVKIEIREGADYKLRVASVNAAGEPPGE	8592
QY	6355	-----RPSMQVTIE---DVQAQGTGTAQFAALIEGDPQSPVTVWKSQVLDVST	6401
Db	8593	TQPVTVAEPOEPPAVELDVSKGIGIMAGKTLRIIPAVVTGRPVTKWTKKEGEL-DKD	8651
QY	6402	RLSQOQEGTTSYLVLRHVASKDAGVVTCLAQNTGGQVLCKABELLVLGGDNEP----DSEK	6457
Db	8652	RVVIDNVGTKSELIIDALRKHGRVITATNSCGSKGFAAARVEVP---DVPGVLDLXP	8708
QY	6458	QSHRRK--LHSFYEVKEEIGRGVFGFKRVQHKGNKILCAAKFIPLRSRTAQAYRERDI	6515

```

Db      8709 VVTRKMLLWSDPDDGGSEITGFI--IERKAKM----- 8743
QY      6516 LAALSHPLVTLGLDQFETKRLTILILELSCSEELDLRYKGVVTEAEVKVYIQQLVEGL 6575
Db      8744 -----HTWRQPIETERS-XCDITGLLEGQ 8766
QY      6576 HYLHSHGVL-----HLDIKPSNLMVHPAREDIKIDFGFAQNIAPAELOFSQYS 6626
Db      8767 EY--KFRVIAKNKFCGPPVIGP--ILAVDP-----LGP 8797
QY      6627 PEFVSPEIIQONPVSEASDIWANGVISYLSITCSPFAGSDRATL-LNVLEGRVSWSSP 6685
Db      8798 P--TSPE-----RLTYT-----ERQSRSTITLDWKEPNSGGSP 8828
QY      6686 MAHLSE-----DAKDFIKATLQAPAPPSAAQCLSHPWFLKMPABEAHFINTKQLKFL 6741
Db      8829 IQGVIIIEKRHKDPF-----ERVNKRLCPTTSFLVNLDEHQMYEFVK----- 8873
QY      6742 LARGRWQSLMSYKSLVMRSIPELLRGPDPSPSLGVARHLCRDTGGSSSSSSSDNELA 6801
Db      8874 -----AVNEIGSEPSLP-LNVV-----IQDDEVP 8897
QY      6802 PPARAK-----SLP--PSPVTHSPL-----LHPRGFLRPSASL---PEAE 6837
Db      8898 PTIKRLSVRGDTIKVKAQEVHIPADVTGLPMFKIEWSKNETVIERPTDALQITKBEVS 8957
QY      6838 ASERSTEAPAPSPGEGAPPAAGCQVPRHSVIRSLFVHQAGESPERGALAPGSRHPR 6897
Db      8958 RSEAKTELSIPKAVREDKGTITVTSNRGLSVFRNVHVEYDR-----PPPRNLAV 9009
QY      6898 RHLLKGGYIAGALP---GLREPLMEHRLVEEAAAREEAQ-----TLIAK-----A 6940
Db      9010 TDIKAESCYLTWDAPLNDGGSE--ITHYVIDKDRASKKAEWEVBVTAVEXRYGIWKLI 9067
QY      6941 PSFETALRLPASGTHLAPCHSHSLHDSPTSPRPSSEACGEAQLPSAPSGGPIRDMGH 7000
Db      9068 PNGQYEFVRVANKY---GISDECKDKVIQDP-----YKLPGP--GKP-KVLAR 9113
QY      7001 PQGSKQLPST-----GGHPTAQ--PERPSPDSP-WGQPAFPCHPKQG-----SAPQ--E 7045
Db      9114 TKGSMVLSWTPLDNGGSPITGVWLEKREESPVKSRVSRAPITKVLKGVFNVPRLLE 9173
QY      7046 GCS--PHFAVACPFGSPGSCKEAPLVPSP--FLGQOAPPAPAKASPPLDSKMGPGDI 7103
Db      9174 GVKYQFRAMAINAAGIGPPS-----PSDPEVAGDPIFPGP-PSCPVEVDKT-KSSI 9224
QY      7104 SLGRKPGPCSPGASQASSOVSLRVGSSQVGTPEPGPSLDAEGW---TQBAEDLSDS 7161
Db      9225 SLGWKP---PAKDG-----SPIKGYIVEMQEGT-----TDWKEVNEPDKLIIT 9266
QY      7162 ----TPTLQRPQBOATWRKPSLGRGGYAGVAGYGTFAFGDAGGMLGQGPMMARJAWAV 7217
Db      9267 CECVVPNLKE-----LRKYAF-----RVK-AV 9287
QY      7218 SOSSEEEQEARAESQEEQOEARAESPLPOVSARVPYEVG-----RAPTRS 7264
Db      9288 NEAGESPSDDTGIPATDQEE-----PEVFI-----DIGAQDLCKVAGSQIRIPAVI 9337
QY      7265 SPEPTP---WEDIQGVSLVQ---IRDSGDAEADTISLDISEVDPAYLNSLDYPI--K 7316
Db      9338 KGRPTKSSWEFDGKAKAMKMGVHDIPEDAQLETAENSVIIIPCKRSHTKYSITAK 9397
QY      7317 YLPFEFMIFRKYKSAQEPSPMAEEELAEFP--EPTWMPGELG----- 7360
Db      9398 NKACQKTANCKVKNVDPGPPDKLVSDITRGSCRLSKWKNPDDGGDRIKGYVIEKRTID 9457
QY      7361 -----PHAGL-----EITESESDVALLAAVAGRKRWSSPSRSLFHPGRLPL 7406
Db      9458 GKAWTKVNPDCGSTTFVVVVDLLSQQYFFRVRAENRFGIGPPVETIQTTFARDP--IYPP 9515
QY      7407 DEPAELGLREVKASV-----EHISRLKGRPEGLEKEG----- 7440

```

Db 9516 DPPIKLIKGLITKNTVHLSWPKPKNDGGSPVTHYIIVECLANDPTGTCKEAMRQCNKRDE 9575  
 QY 7441 -----PPRKPGGLASFRSLGKLSW-----DRAPTF-LRELSDETIVL 7476  
 Db 9576 ELQFTVEDLVEGGEYFRVKAANAAGVSKPSATVGPCQCPDMPSPSIDLKEFME--VEE 9633  
 QY 7477 GOSVTLACQVSAQPAQAATWSKDGAPLESSSRVLISATLKNFQLL--TILVVVAE---D 7530  
 Db 9634 GINNVIVAKIKGVPPPTITWFKAPPKPDNKEPVDYDTHVNLVVDCTTILVQSRSD 9693  
 QY 7531 LGVYTCVSNALGVTITTTGVLRKAERSSSPCPDIDG-----EVYADGVLLWKVPVSYGP 7585  
 Db 9694 TGLYITITAVNLGTASKEMRLNLVLRPG---PPVGPPIKFESVSADQMTLSWFFPKDDGG 9749  
 QY 7586 ---VTYIVQ-CSLEGGSWTTLASDIFDCYLTSKLSRGTYTFRCTACVSKAGMG-PYSSP 7640  
 Db 9750 SKITNYVIEKREANRKTWVHVSSEPKCTYTIPLLEGHEVYFRIMAKNQYIGIEPLDSE 9809  
 QY 7641 SE 7642  
 Db 9810 PE 9811

RESULT 2  
 I38346  
 elastic titin - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 21-Jul-2000  
 A:Accession: I38346  
 R:Labeit, S.; Kolmerer, B.  
 Science 270, 293-296, 1995  
 A:Title: Titins: Giant proteins in charge of muscle ultrastructure and elasticity.  
 A:Reference number: A57430; MUID:96026330; PMID:7569978  
 A:Accession: I38346  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-7962 <RES>  
 A:Cross-references: EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017427  
 C:Genetics:  
 A:Gene: GDB:TTN  
 A:Cross-references: GDB:127867; OMIM:188840  
 A:Map position: 2q31-2q31

Query Match 8.0%; Score 3313; DB 2; Length 7962;  
 Best Local Similarity 21.5%; Pred. No. 8.6e-90;  
 Matches 1710; Conservative 1105; Mismatches 3287; Indels 1868; Gaps 306;

QY 10 PRFLTRPKAFVVSVGKDATLSQIVGNPTPOVSWEKDOQPVTAGARFRLAQDGLYRLTI 69  
 Db 10 PAIITPLOTVTSEGOAPAFQCRVSGTDL-KVSHYKDKKIKPSFFRMTQFEDTYQLEI 68  
 QY 70 LDALGDSGVYVCRARNAIGFAAAGLVQVDAEACABQ-----APHFLRPT 117  
 Db 69 AEAYPEDSGTTFVANNVAVGVSSSTANLSLEAPESILHERIEQIEMEMKAAPIVKRKIE 128  
 QY 118 SIRVREGSEANFRVCGGSPRAVSWSKDGRRLGEPDGPVRVVEELGEASALRRAARPR 177  
 Db 129 PLEVALGHAKPTECIQAPNWRQWFKAGRIEYSDKCSIRSKY--ISSLEILRTQVV 186  
 QY 178 DGGTYVEAENPLGAASAAAALVVDSDAADTASRPGTSTAALLAHLORREARMAEGAPA 237  
 Db 187 DCGEYTCASNEYSVSVCTATLV-----TEA 213  
 QY 238 SPSPSTGTCTCTVTE--GKHARLSCVTGPKEPTEWKKDQGLVTEGRRHVHYVEDAQENFV 295  
 Db 214 YPTFLSPKSLTTFVGAARAFICVTGTFTVETIWKDGAALSFSNNWRI-SDAENKHI 272  
 QY 296 LKILFKGSDRGLTCTXTASNLVG-QTYSSSLVVVREPAVPFKRLQDLVREKESATFLC 354  
 Db 273 LELSNTLIQDRGVYSCASKNKFGADICQAEIILIDKP--HFKELEPVQSAINKKVHLEC 330  
 QY 355 EVPQP-STEAARWFEETFLWASAKYIEEGTEERLTVRNVSAODDAVYICETPE---GS 410









Db 4263 TTGDTCTLECTVA-GTPELSTKWFKDGKEL-TSDNKYKISFFNKVSGLKILINAVPSDSGV 4320  
Qy 4504 VTFACRDV---ASALTVLGLDPPEDA-----EVAHSHHTVTLW 4543  
Db 4321 YSFEVQPVKGLSDCTASLQVSDRTVPPSFRKLKTNGLSGSSVWMECKYVGPPIVSVM 4380  
Qy 4544 AAPMSDGGGLCGYRVEVKEGATGOMRLCHELVPGPECVVDGLAPGETVFRVAAGVPV 4603  
Db 4381 ---FHEN-----BISGRKYQTLTNTWC---ALTVMLEBSDSGDYTCIATNMAG 4426  
Qy 4604 AGEVHLFQTVRLAEPKPPPOPSAPESQVAAAGEDVSLELVVAEAG-EVIWHKMER 4662  
Db 4427 SDE-CSAFLTVR-EPPSFV---QKDDPMDVLTGNTVFTSVKGTPTPSVSWFKGSS 4479  
Qy 4663 IQPGREVVSQGOQLMVLKGTFAEDQGEYHCGLAGSGICPAAAFQVALSPASVDEAP 4722  
Db 4480 LVPGDRCNVLESDVAELELFDVDTOSGEYTCIVS-----NEAG 4519  
Qy 4723 QPSLPPAAQEGDLHLHWEALAR---KRRMSREPTLDSISELPEDEGRSQRLOPQAEVAP 4780  
Db 4520 KASC-----TTHLYIKAPAEVKE-----LNDYSIEKKG-----P 4549  
Qy 4781 DLSEGYTADELARTGADLSHTSSDDESAGTPSLVTYLLKAGRPOTSPASKVGPAA 4840  
Db 4550 LILEGTFT-----GTPIISVTWKXG----- 4570  
Qy 4841 PSVKPQQQBEPLAAVRPLGLDSTKDLGDPMSDKAAVKIOAFKGYKVRKEMKQCGPMF 4900  
Db 4571 INVTSPORCNITTEKSPILIPSTVEDAGQNCYIE-NASGKD-SCSAQIILIPPPY 4628  
Qy 4901 SHTEGDEAQVGDALRLCEVVASKADYRARKLKGVEL---TGRHHHIDQLGDCGTCSL 4957  
Db 4629 VKOLEPVKVSVDGSASLQCLAGTPEIGVSWYKGDTKLRPTTYKMH---FRNNVATLV 4684  
Qy 4958 IAGLDRADAGYTCQVKNKGQVTHSACVVVSGSEASESSGELDDAFRAARLRL 5017  
Db 4685 FNQVDINDSGEYIKAKNSVGEVSEASTFLTVOEQ-----KLPSPSRQLRDVQE- 4733  
Qy 5018 PRTKSPAEVSEDEFLSADGPAEPPEPADQTYREDEHFICIRFEALTEARQAVTRFOE 5077  
Db 4734 -----TVGLPVVFDCAISG-----SEPIVSWSYKDG-----KPLKDSNPVQTSFLD 4774  
Qy 5078 MFATLGI-GVEIKLVEQPRVEMCISKETPAVPVPEPPLSLTSDAAVFLTELONQE 5136  
Db 4775 NTATLNFKTRSLAQ-----YSCA-----TNPIGS---ASSARLILTEGNKP 4818  
Qy 5137 VQD-----GVPSVDFCVTQGPMPVSRWFKDGKLEEDDHYMINEDQOQHOLII 5186  
Db 4819 FFDIARLAPVDVAVGESADFECVHTGTQPIKVSNAKDSREIRSGKYQISYLENSAH-LTV 4877  
Qy 5187 TAVPADMGVYRCLAENSMGVSSTKAELRLVDLTSTDYDTAADATESSSYFSAQYLSRE 5246  
Db 4878 LKVDKDSGOYTCYAVNEVGKDSCTAQLNI----- 4909  
Qy 5247 QEGTESITDEGOLPOVVEELRLQVAPGTRLAKQL--KVKGYPAPLXWFKDQOQLTAS 5304  
Db 4910 RLIPPSFTK--RLSTVBE-----TEGNSFKLGRVAGSQPIVAVYKNNIEIQPT 4958  
Qy 5305 AHIRMTGKILHLEIISVTRSDSQYAAYSNANGAAYSSARLLVRGPDEPEKPSADV 5364  
Db 4959 SNCEITFNNTLVLQVRKAGMNDAGLYTCVKSNDAGSALCTSSIVIK---EPKPPVFDQ 5015  
Qy 5365 HBQLVPPMLERFTPKYKKGSSITFSYKVEGRPVPTVHMLREAEERGLVMIGP-DTPEG 5423  
Db 5016 H-----LTPVTYSEGEYVQLSCHVQSGSEPIRIONLKGABE-----IKPSDRCSF 5059  
Qy 5424 TVASSAQOHSVLVLDVGRHQHTYTCIASNAGQALCSASHVSGLPKVEQEKEALI 5483  
Db 5060 SPASGTA--VLELRDVAKADSGDYCKASNVAGSDTTKSKVTIKDKPAVATKAAVDG 5117  
Qy 5484 STFLOGTTQAIASQGLEITASP-ADLGQORKKEPLA-AKEALGHLSLAEVGTBEFLQKLTS 5541  
Db 5118 RLFFVSEPGSIRVVEKTTATFTAKVGG-----DPIENVKWTGKWRQLNQGRVFIHQ--- 5170

Qy 5542 QITEMVSAKITQAKLOVPGGSDSDSKTPS-----ASPRHGRSRPSSIO-----ESSSES 5592  
Db 5171 -----KGDEAKLEI-----RDTIKTDSGLYRCVAFNEHGEIESNVNLQVDERKQGEK 5217  
Qy 5593 EDGDARGEIPIYVYVYADYLPFGABEQDAITLREQQYVEVLDAAHPLRMLVTRTKPTKSSPS 5652  
Db 5218 IEGDURA-----MLKKTPIILKKGAGEE---EEDIMELLKNVDP----- 5253  
Qy 5653 RCGWSPAYLDRRLKLSPEWGAAPPEPPGAEVSEDEYKARLSSVIOELLSSSEQAF--VE 5710  
Db 5254 -----KEYEKYARMYG-ITDFRGLLOAFELK 5279  
Qy 5711 ELQFLQSHHLOHLERCPHVPIAVAGQKAVIFRNVRDIGRPHSSFLQELQOCDTDDDVAMC 5770  
Db 5280 QSOBEETHRLE-----IEETERGERD----- 5300  
Qy 5771 FTIKQAAFEQYLEFLVGRVQABSVVSTAIOEFYKYAEABALLAGDPSQPPPLQHLYLE 5830  
Db 5301 -----EKEFEELVSFQQEL-----SOTEP----- 5320  
Qy 5831 QPVERVQVQALLKELIRKARNRONCALLQEAQAVVVSALPQRAENKLHVSIMENYPGTL 5890  
Db 5321 -----VTLIKD-IENQIVLKNDVAFE-----IDIKINYP-- 5349  
Qy 5891 EALGEPIRQHFIVWEGAPGAMPWKHNHVLPFRNLHVICKPRRSDRTDVTSVFRNM 5950  
Db 5350 ----- 5350  
Qy 5951 MKLSSIDLNDQVEGDDRAFEVWQEREDSVRYKYLLOARTAIKSSW-VKEICIQOORLALP 6009  
Db 5351 IKLSWYKTEKLEPSDK-FEISIDGD---RHLRVKNCOLKQGNRYLVCOPHILASAKL 5405  
Qy 6010 VWRPDPFEBELADCTAELGETVKLACRVY---GTPKPVISWYKDGAKVQVDPHHILIEDP 6066  
Db 5406 TVIEPAWERHLQDVILKEGQT---CTMTVQPSVFNKSEWFRNGRILKPOGRH-KTEVE 5460  
Qy 6067 DQSCALILDSLTVGDSQGYMCPAASAGNCSTLKLIVQVPP-RFNVKVRASPFVEGEDA 6125  
Db 5461 HKVHKLTIAVRAEDQGOYTC---KYBLETSAELRIEABEPIQTKRIONIVVSEHQSA 5516  
Qy 6126 QFTCTIEGAPYQIIRWYKDGALLTGNKFQTLSEPRSGILLVIRAAKEDLGLYCECLV 6185  
Db 5517 TECEV-SFDDAIVWYKGTPELTESQKYNFRNDGRCH--YMTIHNVTPDDEGVY--SVI 5571  
Qy 6186 NRL---GSARASAEIRIQSPMLQAOBQCHREQLVAVEDTTLERADQEVTSVLKLLGPK 6242  
Db 5572 ARLEPRGEARSTAEIYL-----TTKEIKLELKP- 5599  
Qy 6243 APGSPGDLTGPGP-CPRGAPALQETGSOPTVTGTSEAPVPPVPPVQPLLHEGPEQPEPA 6301  
Db 5600 ---PDIPDSRVIPTPMPTRAVPPEI---PPVV---APPVPLLPTP-----EEKPP- 5643  
Qy 6302 IARAQEWTPVIRMEGAAMPFAGCTGELLMDVHSHVRETTQRTYTYQOIDTHARTAPPSMOV 6361  
Db 5644 -----PKRIE-----VTKKAVKOAKKV-----VAKPREM-- 5668  
Qy 6362 TIEDVQAOGTGAQFAIIEGDPQPSVTWYKDSVQLVDSTRLSQOQEGTYSILVLR--HV 6419  
Db 5669 -----TPREIIVKXPPPTLLIPAKAPEIID---VSSKAEVFKWITIRKKEV 5713  
Qy 6420 ASKDAGVYTCIAQNTGGQVLCKAEALIVLGGDNEP--DSEKQSHRRKLHS-FYEVKEEIGR 6476  
Db 5714 QKEKEAVY-----EKKQAVHKEKRVFIESFEPEYDELEVEPYTEPFEPQYVEEDDEYE 5767  
Qy 6477 GVFGFVKRVQHKNKILCAAKFIPLRSRTAQAAYERDILAAALSHPLVTGLDQFTRKT 6536  
Db 5768 EIKVEAKKEVHEWE-----EDFEGGEFYERE-----EGYD----- 5799  
Qy 6537 LILILECSSEULLSLRYKGVVTEAEVKYIQQOLVEGLHYLHSH-----GVILHLDIKP 6590  
Db 5800 -----EGEEWEAYQEREVIQVKEVYEE-----SHERKVPKVPKPKAPP 5841

QY	6591	SNILMVHPAREDIKICIDGFGPAQNTIPABLOQSVQSGSEFFVSPEIIQONPVSEASDIWAMG	6655
Db	5842	PKVIKKVIVIEKIE-----KTSRMEEEKKVQVTKV--PE-VSKKTVPOKP-----	5883
QY	6651	VISYLSLTCSSFPFAGESDRATLLNVLEGRVSSWSPMAAHLSEDAKDFIKATLQAPQARP	6710
Db	5884	-----SRTPVQEE-----VIEKVPVAHTKKXVISEEKKPFASHTSEEVSVTP	5927
QY	6711	SAACLSHPWFLKMPABEAHFINTKQLFLARSWRQSLSMYKSILVWRSIPELLRGP	6770
Db	5928	EV-----QKBIVTBEEKIHV--AVSKRVEPP-----	5950
QY	6771	PDSPSLGVARHLCRDTGSSSSSSSSDNEIAPFARAKSL-PPSPVT-----	6815
Db	5951	PKVPFL-----PEKPAEEVAPVPIPKVPEPPAKPVEVPKKVPVPEKKP	5995
QY	6816	-----HSPLLHPROFLRPSASLPEEASEASERSTEAAPPASPEGAGPPAAQCVGPHRSVI	6870
Db	5996	VPVKKGFPAAPKVPVEPKKFPVPEKEITFVPVAKKKEAPAKV-----PEVQKGVVTEKI-	6050
QY	6871	RSLFYHOAGESPEHGA-----LAPGRRHPARRRHLLKGYIAGALPGUREPLMEHRV-	6923
Db	6051	--TIVTQEEESPFAVPEIKPKKVPEERKVPVPRKEEVPPPKVPALP--KKVPVPEKVA	6106
QY	6924	-----LSEEAAREHQATLLAKASFETALRLPASGTHLAPGH--SHSLEH	6966
Db	6107	VPVPVAKKAPPRAEVSKTVVEEKRETFVABEKLFSFAVPQRVETRHEVSAEESWSYSEE	6166
QY	6967	DSPS-----TPRSSEACGE-----AORLPSAPSGCAPIRDWIG	6999
Db	6167	EGVLSIVYBREERESEEAETVEYEVWEEPEEYVVEEKHLISKRVAEAP-----EVT	6220
QY	7000	HPQSGK-----QLPSTGGHFGTAQ--PERPSPDSMGQ--PAPFCHPKGSGAPQEGCSHPHA	7052
Db	6221	ERQKKIIVLPKIEPAKITEEPPPAKVPPEAPKPIVPEKKVPAPV--PKKEKVPVPKVPPEEK	6278
QY	7053	VAPCPPGSGPPGCK-EAPL-----VPSPPFLGQOPAPPAPAKASPLDSKMG-	7099
Db	6279	-KPVPKEKVPVKIKMEPIPAKVTEKHMQITQEEKVLVATVTKEAPPPKARVDEEPRAV	6337
QY	7100	PGDISLPGRPKPGPCSPGSGASQSSVSLRVGSSVGTGTERPSLDAEGMTQE--ABD	7157
Db	6338	PEEKVLKLPKREB-EPPAKVTBFRKRVVKEEKVSIAPKREPOPIKEVTIMEEKRAV	6396
QY	7158	LSDSPTPLQRPQEQATWKRKSLGGRGGYAGVAGTFAFGDAGGMLGQPMWARIAWAV	7217
Db	6397	LEEEAVSQREEEVEEYDYKEFEYEYETEYDQY-----	6433
QY	7218	SQSEEEQCEBARAFSQSEEQCEARASPLPQVSARVPVEVGRAP-----TRSSPE---PT	7269
Db	6434	--EYEBERYEYEBEYIETEKPIPVKVPPEEPVTPKPAAPPKVLKXAVPEKVPV	6491
QY	7270	PWEDIGQVSLVQIRDLTSGDAEAAATISLDISEVDPAYLNLSLDIYDKYLPFEFMIFRPV	7329
Db	6492	P-----IP	6494
QY	7330	KSAQPEPPSPMABEEAABFPETWPMFGLPGHAGL-----EITESEEDV-----DALLA	7379
Db	6495	KGLKPPPPK-VPBEPKKVPFEKI-----HISITKREKQVTEPAAKVPMKPKRVVA	6544
QY	7380	EAAVGRKRKWSRSRSLFHPFGR--HLPLD-EBAELGLRERVKASVEHLSRL-----	7429
Db	6545	EKKVPVRKEVAP-----PVRVPEVPKELEPEEVAFEFV---VTHVEEYLVBEERY	6594
QY	7430	-----KGRPEGLEKEGPPRKKP	7446
Db	6595	IHEEEEFITEEVVVPIPVKVPVPRKVPPEKKPVPVPKKKEAPPAKV	6644

C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T42633  
R:Yajima, H.; Ohtsuka, H.; Kawamura, Y.; Kume, H.; Murayama, T.; Abe, H.; Kimura, S.; Ma  
Biochem. Biophys. Res. Commun. 223, 160-164, 1996  
A:Title: A 11.5-Kb 5'-terminal cDNA sequence of chicken breast muscle connectin/titin re  
A:Reference number: Z22221; MUID:96254045; PMID:8660363  
A:Accession: T42633  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-4162 <1A>  
A:Cross-references: EMBL:D83990; NID:G1513029; PIDN:BAAL1908.1; PID:G1513030  
A:Experimental source: breast muscle  
C:Keywords: skeletal muscle

Query Match	6.4%;	Score 2629;	DB 2;	Length 4162;
Best Local Similarity	23.4%;	Pred. No. 6e-70;		
Matches 1110;	Conservative 694;	Mismatches 1874;	Indels 1074;	Gaps 172
Qy	9	APRFTRPKAFVSV-GKDATLSQIVGNPTQVSWEKDQPVTA----	GARFRLAQDGD	63
		:	:	
		:	:	
		:	:	
Db	5	APTF-TQPLQSVVLEGSAATFEAHISGFVPVSWYRDQVLSAATLPGVQISFS-DGR	62	
		:	:	
		:	:	
		:	:	
Qy	64	LYRLTILDALDSCQVYCRNAILGAFAAVGIQVDAABACAEQAPHELLRPTSIRVRE	123	
		:	:	
		:	:	
		:	:	
Db	63	A-KLVIPSVTEANSGRYTIQTNGSGGATSTAEILLTAGTA----	PPNFSQRLQSMATQ	117
		:	:	
		:	:	
		:	:	
Qy	124	GSEATFCRCVGSRPVAVSWSKDGRRLGPDGPRVRVEELGEASALIRIAAARPDGGTYE	183	
		:	:	
		:	:	
		:	:	
Db	118	GSQRLDVRVTGIPTEPVVKFYDGVET--QSSPDFQLQEGDLYSLIIAEAYPEDSGTYS	175	
		:	:	
		:	:	
		:	:	
Qy	184	VTAENPLGAAGAAALVDSADATAGRPOTSTAAILLHORREA-----	229	
		:	:	
		:	:	
		:	:	
Db	176	VNATNNVGRATSTAEILLIQEEEAAPPAK-TKTIIVSTAQISQTRQARIKKIETHFDARS	234	
		:	:	
		:	:	
		:	:	
Qy	230	-----WRAEGAPA-----SPPSGTGTCVTVEGKHARLSCYVTGEPKPEVTWKKDGLV	278	
		:	:	
		:	:	
		:	:	
Db	235	LTSVENWTEGAAAOQLPHKAPPPMPRRP-----TSKSPTFPVITAK-AQWA	279	
		:	:	
		:	:	
		:	:	

QY 787 YQLSVQGLARFLHKDMAGSCVDVAGGPAOF-----ECETSEAHVHVHYKDGMLGHSGE 842  
Db 622 -----ESAPPQFPFTEAAETVKAHYDVETKKE--VDVSIKGE 656  
QY 843 RFLQEDVGTTRHLVAAVTVTRODEG--TYSRCVGEDSVDFRLRVSEPKVFAKSQLARKL 900  
Db 657 A-VRED-----HLL-----LRKESEAKVTETARV--PVPABIPVTPPLVWG-----LKNKT 700  
QY 901 QAEAGASATLSCEVA-QAQTETVTKYDKGKLSSSSKVCMBATGCTRRLLVQQAGQADAGE 959  
Db 701 VTE-GESVTLTCHTSGHPQPTVWYREDYKLESSMDFQITFKAGLAELVIREAFEDSGR 759  
QY 960 YSC-----EAGQRLSFHLVDVKEPKVFAKQVHAHSEVQAEAGANATLSCEVAQAQAEVW 1015  
Db 760 FTCTATNKAGSVSTSLHVK-----VSEETETRETIS-----792  
QY 1016 YKDGKLSLSLVKVAEAKGCRRLVVOQAGTKDAGDYSCAARGORVFRHLHITEPKMFA 1075  
Db 793 -----EKVTEKSVETKD-----VME-----DVSAAEVSEGPV-----PPFFI 830  
QY 1076 KEQSVHNEVQAEAGASAMLSCEV-AQAQTEVTWYKDGKLSSSSKVGMVEVKGCTR--RLV 1132  
Db 831 RKPVVHKLIE---GGSIIIFECQVGNPKPHVLMKGGVPLTTGYRYKVSRYKRETGECKLE 887  
QY 1133 LPQAGKADAGEYSC-----EAGG-----ORVSHLHIT-----EPK- 1163  
Db 888 ISMTFADDAGSYTIVIRNKPGEASATVSLLEEADYEAIIKQSQEMMTQTQVAYVQEPKV 947  
QY 1164 -----GVFAKEQSVHNEVQAEAGTTAMLSCEVAQPOTEVTWYKDGKLSSSSKVRM 1214  
Db 948 AEVAPPISYGDFDEYEKEQ-----ALIRKKMAKDTVMVTFVEDEEFHISSPEER 998  
QY 1215 EVKCTRLVVOQVGKADAGAYSCEBAGQORVSPQLHITEKAVFAKQLVHNEVRETAG 1274  
Db 999 LIKSEIURLII-----KTLDELLEEDGEE---NMIDISEAIGAGFDLRLKNRYTEGT 1050  
QY 1275 SATLSCE-VAQAQTEVTWYKDGKLSSSSKVRIEAAAGMRLVVOQAGQA-----1323  
Db 1051 QVTFCKTGTGPLKIAWYKDGKRIHGERVHME-----VLQGSASLRLPVVLE 1101  
QY 1324 DAGBYTCEA-----GGQRLSFHLVSEPKAVFAKE-----1353  
Db 1102 DEGIYVTPASNMKNNAICSAKLYV-EPVAPTATPGVMPGPEVMRRYRSISPRSPSRSPAR 1160  
QY 1354 -----QLAHRKVQAEAGAIATL-----SCEVAQAO-----EVTWYK 1385  
Db 1161 SSPSCSPARRLDETDGQLERLYKPVFLKPTSVKCSGGQTARFDLKVGRPMPETWFFH 1220  
QY 1386 DGKLS--SSKVRMEAVGCTRRLLVVOQACQADTGEYSC-----BAGQRLSFSLDVAEPK 1439  
Db 1221 NGQVVNDYTHKIVIKEDG-TQSLIIVPAMPEDSGEWAVIAQNRAKASVSVTLSEB---1276  
QY 1440 VVFAKEQPVH-----REVQAQAGASTLSCE-VAQAQTEVMYKDGKKL--SFSSKVR 1489  
Db 1277 ---AKEDILVRPFVERLNVSVKGSRLHMAVKATGNPNPDIVLWKNSDIIVPHKYPRIR 1333  
QY 1490 MEAVGCTRRLLVVOQAGADAGEYSCAGSQ-----RLSFHLHVAE-----PKA 1532  
Db 1334 LEGTKGAALNIESTARQDAWYATATAINKAGROTTRCKNVNVEHAEPERELIIPKG 1393  
QY 1533 VFAKEQPASREV-----QAEAG-----TS-----ATLSCE 1557  
Db 1394 TYKAKEIAAPELEPHLRYGQEQWEEGLDYKEXQKQPFPPKKULTLRLKQFGPAHFECR 1453  
QY 1558 ---VAQAQTEVTWYKDGKLSSSSKVRM--BAVGCTRRLLVVOEAGQADAGEYSCKA---1608  
Db 1454 LTPIGDPTVMVWELHDGFLPBAANRLRMINEFGYCS--LDYGVAYSRRDSGVITCRATNKY 1511  
QY 1609 GDQRLSFHLHVAEPKVVFAKEQ-PAHR-----EVQAEAGA-----1642  
Db 1512 GTDHTSATLIVKDEKSLVEESQLPEGRRCMQRIEELERMAHEGALPAVAVDQKEKQKPEL 1571

QY 1643 -----SATLSCEVAQAQTEVT-----NYKDGKLSSSSKVRVAEAVGCTRRLLVVOQA 1688  
Db 1572 VLVPPARVLEGETARFRCTVGTGYPLPKVNWYNSQLIRKSKRFLRYDVG-IHYLDIVDC 1630  
QY 1689 GOADAGEYSCAAGGQRLSRLHVAELPQISB-----RPCRRLPLVVKEHEDIL 1738  
Db 1631 KSYDTGEVUKVTAENPE-GFIEHKVLEIQOREDFRSLVRAPEP-RHPFVTEPCKLLE 1688  
QY 1739 TATLATPSAATYTWLKGVEIRRSKR--HETASQGDTHLTLVHGAQVLDLSAIIYSCRVGAE 1796  
Db 1689 VQKIDKPABATT---KEVVKLKRAERITHEKLSE-----1719  
QY 1797 GQDFPVQVBEVAAKFCRLLEPVCGLGGTVTLACELSPACAEVWVRGCGNTQPRVGRKQPM 1856  
Db 1720 -----ESELRSKPKRRT-----1733  
QY 1857 VAEGPVRSLTVLGLRA---EDAGEYVCESEDD--HTSAQL-----TVSVP-- 1896  
Db 1734 --EGYVEA-ITAVELSRKXKDESEBMLKKTKEELLHWTKEIPEEBKKALPPEBKITIPTF 1791  
QY 1897 -----RVVFMESGLSTVVAEBEGEATFQC-VVSPSDVAVVWRDGGALLPSEK- 1943  
Db 1792 KPEKVELSPMBAPKIFERIQSQTVAQSDAHFRVVRVVGKPDPECFQWFRNGVOIERTDI 1851  
QY 1944 FAISQSGASHSLTISDLVLEDAGQITVEA---EGASSSAALFVRBAPVL--FKKLEPOT 1998  
Db 1852 YWYWPEDNVCELVRDVTADDSDASIMVKAIVNIAGETSSHAFLLVQAKOLISIQNLQDVV 1911  
QY 1999 VVERSSV-TLEVLTTRPPELWRWTNATALAPKKNVEIHABGARHLVLHNVGPADRGFF 2057  
Db 1912 AKERDSMATFECETSEPFIFKVKFKNGIEIHSGEKYRMSDKRAHFLSVLAVEMSDADDY 1971  
QY 2058 GCETPDD---KTOAKLTIVEMROVRLVRLGLOVAEAREQGTATMEVOLSHADVDGSTRDGL 2114  
Db 1972 SCALVEDESVKTKALIVEGAVVEFIKELEDVEVEPESFTGELECEVSEPEDIEGKWHGV 2031  
QY 2115 RFQOQCTCLAVRGMHTLTLSLRPEDSGLMVFAKGVHTSARLVVTELPSVFSRPLQD 2174  
Db 2032 ELSSNNKYVLASRRGRRIITIKDVKNDQOGEYSFVVVDGKRTCKLKMKPRMPTILQGLTD 2091  
QY 2175 VVTEKEKVTLSCELSRPNVDVRLWKDGVLELRAGKMTAAQAGACESLTIYRCEPADQGV 2234  
Db 2092 QKVCEGDIYQLEEVKVSVENVEGVWMDG-----2119  
QY 2235 YVCDADHAQSSASVKV--QGRTYTLIYRRLVLAEDAGEIQFVAENAESRAQLRVKSLPVTL 2292  
Db 2120 -----HEIQSSDRIHVLQKQSHMLLIEDATQEDSGTYSFSIPGLESLSTGTQVTVSVBEI 2174  
QY 2293 VRLPRDKIANEXHGRGVLECOVSRAS-AQVRWFKGQOELOPGPKYELVSDGLYKRLIISDV 2351  
Db 2175 IVPLKDVHVVEGTKAILECKVSAPDVTSSKWYLNHDQIKPDERVQAVCKGTQRLVITRT 2234  
QY 2352 HAEDEDTYTCADAGDVKTSAQFFVEBQSIIVIRGLQDVTWMEPAPAFECETSIPIVSRPPK 2411  
Db 2235 HASDEGHYKLVGKVETSQNVTVBE--TEIIRGLHDICTETQNVSPFVELSHSGI-DVI 2291  
QY 2412 WLIGKTVLOAGNVGLEQGTVHRLMLRRTCTMTGPVHFTVVGKSRSSARLVSDIPVVL 2471  
Db 2292 WHFKGQEIKAQPKYKIEARGKIYKLTIVVQWMDDEGEYVYFAGGKKTSGKLIVAG--GA1 2349  
QY 2472 TRPLEBKTRELQSVLSCDFRPAKAVQWYKDDTPLSPSEKFKMSLEGQMAELRLRLM 2531  
Db 2350 SKPLADLTVAESQRAVFECEVANPSEGOVLKNGKPLPMDIQTRAETDGVKRLNIPAAK 2409  
QY 2532 PADAGYVRQAGSAHSSTVTVAREVTVTGPLQDAEATEGWAFCWAGSFCSELSHED-EVEW 2590  
Db 2410 MDDWGEYSVEIASSTSAKLHVEAVKIKT--LKNLTVTETQBAVSPVSELSPDVKGALW 2467  
QY 2591 SLNGMPLVNDSPHEISHKGRHTLVLSIQORADAGIVRASSLKVSALEVRVFKVPL 2650  
Db 2468 IKNGVELENDKYEISVKGTVHTLKIHKCVWTDSEVYSFKLGIKANABL--HVEITVKII 2525  
QY 2651 KALDDLASABERGTLAQCEVSDPEAHVWVRKQGVQGLPSDKYDFLHTAGTRGLVWHDVSP 2710

Db 2526 KKPXDVTALNANVVSFELSVSHDTPVVRWPHKVNVELKQSDKYKMI SORKVHKLMLHNISP 2585  
 QY 2711 EDAGLYTCHVGSEETRAVRVHDHVGITKRLKTMELVEGESSCFEVLHSHESASDAMW 2770  
 Db 2586 ADAGEYAFVQGLCKAKLFVETIH--ITKMKSEIEPETKTASFQCEVSHFNV--PSVW 2641  
 QY 2771 TVGGKTVGSSRFQATROGRKYIILVREAPSDAGEVVSFVRLGTSKASLIVREPAII 2830  
 Db 2642 LKNGVEIEMSKFPIVVOGKLHQLNIMNTSEDSAEYTFVCGNDRVSATLTV--KPLIIT 2699  
 QY 2831 KPLEQWVAPGEDVELRCELSRAGTPVHWLKORKAIRKSKYDVVCGTWMAMLVIRGASL 2890  
 Db 2700 SMLEDINAEKXDTITFEVTNYEGISYKWLKNGVEIKSTDKCQIRTKKLTHLSIRNVHF 2759  
 QY 2891 KDAGEYTCVEASKSTASLHVVEKANGFTEELTNLOVEEKGTAFTCKTEHPAATVTRK 2950  
 Db 2760 GDAEYTFVAGKAASSATLYVEARHIEFRHKIDIKVEKRAIFCEIESEPDVQVQMMK 2819  
 QY 2951 GLLERASGRKHPQSEGLTLELTISALEKXADSDTYTCDIGQASRAQLLVQGRVHIIE 3010  
 Db 2820 DQELQIGDRMKIQREKYVHRELIIPSTKMSDAGQYTVVAGNTSSANLIVEGRDVRIRSI 2879  
 QY 3011 LEDVDQEGSSATFRCRISPANYEPVHFLDKTFLHANELNEIDAQPGY-----HVL 3064  
 Db 2880 RKEIQVIERQBAIEFEVNEDDIEP-QWYKD-----GIEINFHYERYSYVVERRHMS 2933  
 QY 3065 LRQALAKDSGTIFYEAGDQASAAALRV--EKPSVFSRELTDTATITEGEDTLVLCETSTC 3122  
 Db 2934 IFETTSYDAGEYTFVAGNRSSVLYVNAPEPQII-QELQPTVSGSKFAECALISGK 2992  
 QY 3123 DIP-MCWTKQKTLRGARCOLSHEHRAQLLITGATLQDSGRYKCPA-----GGACSSSI 3177  
 Db 2993 POPKVSUYKDDQQLSPGCKFLHDAQEYTLITETFPEDSAVYTCBAKNYGVATTSAS 3052  
 QY 3178 VRHARPVRFQ-----ALKDLEVLGGATLRCVLSSVAAPVKWCYGNVLR 3225  
 Db 3053 LSVIPEVVSPELVPVPPVPAVIVPLRDVATSEQASRFQCVTGTDLKVSWSYKDBREIK 3112  
 QY 3226 PGDKYSLRQEGAMELVVRNLRPODSGRY-----SCSFGDQTTSATLTVTALPAQFTGKLR 3281  
 Db 3113 PSRFRMTQFBDTYQLBIAEAYPDEGTYTFVASNSVGQVTSAILKLEAPEKIMYEKLE 3172  
 QY 3282 NK-----EATGATATLRCELSKTAPE--EWRKSETLRRGDRYCLR 3321  
 Db 3173 BEIEMEVKAPILRRRLLEPVAHVNHAKFTCEVETTPNVKFWYKAGREIYDQKYSIR 3232  
 QY 3322 QDGMCELQIRGLAMVDAEYSCVCGBERTSAS-----LTIRPMPAHFIFGLRHQESIEGA 3377  
 Db 3233 SSNYLSTLEIPRPQVDCGEYSCKASQHSVSSTAFLTVE-PPRFIKLDDSSRLVKQH 3291  
 QY 3378 TAT-LRCEL--SKAAPVWRKGRSLRQDGRHSRLQDGAECLEIQGLAVADAGEYSC-- 3432  
 Db 3292 DSTRYECKVGGSPKIKVTYKGETEIHPSKYSMSFVDSVAVLEMHNLVSDESDGYSCBA 3351  
 QY 3433 --VCGBERTSATLTVKALPAKFTGLENEBAVSGATAMLMCELSKVAP--VEWRKGPENL 3488  
 Db 3352 QNPAGSASTSTSLVKAPPA-FTKKPHPVQTLKGSVDHLECELOGTPPQISWYKDKREI 3410  
 QY 3489 RDGRYILRQEGTCELEIQGLAMADAGEYLCVC-----GERTSATLITRALPARIEDV 3544  
 Db 3411 RSSKKYKMGSENYLASIHLNVTADYGEYHCKAVNDVSDSGISGVTLLRA-PPTFVKKL 3469  
 QY 3545 KQEARSGATAVLQCELSNAAPVE--WRKG-SETLRDGRYSLRQDGTCKELQIRGLAWA 3601  
 Db 3470 SDVTVVVYGEIIELOAAVEGQPTISVLWLKDKGIIRESLENLWISYSENVASLIGNAET 3529  
 QY 3602 DTGYSVCVCGERTSAMLTVRALPIKFTETGLRNEEATGATAVLRCELSQAPVEMWKGH 3661  
 Db 3530 NAGKYIC----- 3536  
 QY 3662 ETLRDGRHSRLRQDGAECLEIQIRGLVAEDAGEYLCMGKERTSAMLTVRAMPSKFTGLR 3721

Db 3537 -----QIK-----NDAGFQECF-----AKLTV-LEPAVIVEKPG 3564  
 QY 3722 NEEATEGDTATLWCEL--SKAAPVWRKGRHETLRDGRHSRLRQDGRSCELOIRGLAVVDA 3779  
 Db 3565 PVKVTAGDSCTLECTVDGTPELTARWPFQDGNELSDTHKYKISFFNKVSGLKLNAGLEDS 3624  
 QY 3780 GEYSC---VCGBERTSATLTV--RALPARIEDVKNQEARREGATAVLQCELSKAAP--V 3831  
 Db 3625 GEYTEVKNSVGKSSCTASLOVSDRIMPSPSTRKLKETYQGLGSSAVLECKVYSGPILV 3684  
 QY 3832 EWRKSETLRGDRYSLRQDGTGRCELOIHLISVADTGEYSC-----VCGERTSATLTVRA 3887  
 Db 3685 SMFHDQBITSGDKYQATLTDNTCSLKVNGLQESDMGTYSCTATNAGSDSCSAFLSVRE 3744  
 QY 3888 POPVF--REPLQSLQAEESGSTATLQCELSBPTATVVMKGLQLOANGREPRELQCGTA 3944  
 Db 3745 P-PSFVKPPEPNVLSGENITFTSI--VKGSPPLEVWFRGSIELAPGHKCNITLQDSVA 3801  
 QY 3945 ELVLQDLQREDTGEYTCGSGQATSATLTV--TAAPVRFRLRELOHQEVDDEGTAHLCC 4001  
 Db 3802 ELELFDVQPLQSGDYTCQVNEAGKISCTTHLFVKPEAKFVMKVDLSVEKG--KNLILE 3859  
 QY 4002 LSRAGA---SVEMWKGSLQLPFCAYQWQDGAABELLVRGVEQEDAGDYTC----DTGH 4054  
 Db 3860 CYTGTPTPISVTWKNQGVILKHSKCSITTTETSAILEIPNSKLEDOQYSCHEIENDSQ 3919  
 QY 4055 TQSMASLSVRVPRPKFTRLQSLQEQETGDIARLCCQLSDAESGAVVQWLKGVELHAGPK 4114  
 Db 3920 DNCHCAITLSP-PVFTPLBPVQVTVGDSASLOQCVAGTPE-MVSWYKGDTKLRGTAT 3977  
 QY 4115 YEMRQCATRELLIHOEAKTGEYACV--TGQKTAASL-----RVTEPEVITVRLV 4166  
 Db 3978 VMHFRNQVATLVFSQVDDSDSGEYICKVENTGAEATSSSLTLYOERKLPPSFT--RKL 4035  
 QY 4167 DAEVTADEDFSCSEVSRAGATGVQWCLQGLPLOSNEVTEVAVRDGRIHT-----LR 4218  
 Db 4036 DVHETVGLPVTFDCGIAGSEPIEVSFKDNVRVK-----EDYNVHTSFIDNVALIQ 4086  
 QY 4219 L-----KGVTPEDAGTVSFHLGNHASSAQLTVRAVEVITL--EPLQDVQLSEGQDASQCR 4272  
 Db 4087 ILKTDKSLMGQYTCASNAIGTASSGKVLTEGKTPFPFDTPTIPVDGIIGESADFECH 4146  
 QY 4273 LSRASGGQEARWA 4284  
 Db 4147 ISGTQPIRVTTWA 4158

## RESULT 4

T29757

protein UNC-89 - Caenorhabditis elegans

C.Species: Caenorhabditis elegans

C.Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 03-Dec-1999

C.Accession: T29757

R.Du, Z.; Le, T.T.; Wilson, R.

submitted to the EMBL Data Library, May 1997

A.Description: The sequence of C. elegans cosmid C09D1.

A.Reference number: Z20679

A.Accession: T29757

A.Status: preliminary; translated from GB/EMBL/DDB

A.Molecule type: DNA

A.Residues: 1-6642 &lt;DUZ&gt;

A.Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN000019; CESP:unc-89

A.Experimental source: strain Bristol N2; clone C09D1

C.Genetics:

A.Gene: CESP:unc-89

A.Map position: 1

A.Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6

/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match

Best Local Similarity 20.8%; Pred. No. 1.3e-65;

Matches 1520; Conservative 972; Mismatches 2737; Indels 2064; Gaps 304;



2284 E--SVKPSAIVTGKPMFNVVTVLNNKKLIQSEEVKVKVHETGKTSIRIQKPLMEHN--- 2338  
2052 ADRGFFGCTP-----DDKQTA-KLTVM--RQVRLVRGLQAVEARE-QGT 2093  
2339 ---GTRVEAENVSGKVQATAQLKVDKTEVPKFTTNWDRQVK-----EGEDVKFT 2387  
2094 ATMEVQLSHADVDSWTRDGLRFOOGPTCHLAVRGPWHTITLSGLRPEDSGLMVFRAG- 2152  
2388 ANVE---GYPPSVAWTLNGEPVSKHFNITVDKDGESHTIISAVTPEQAGE-SCEATNP 2444  
2153 VHTSARLV-----VTELPVSFSRPLQDVVTEKEKVTLECELS--RPNVDVRLWKDGVE 2204  
2445 VGSKRQVQLAVKKVGDP-TFAKNLEDRLLITEGELTLMDAKUNIVKPKKITWLKDGVE 2503  
2205 LRA-GKTMIAIAQAGACSLTIYRCFADQGVVCDAMD-----AQSSASVKVQGRTYTLY 2259  
2504 ITSDGHYKIVEEDGSLKLSILOTKLEDKGRITIKAESFQVABCSASLV-----VK 2556  
2260 RRVLAEDAGEIQFVAENASRAQLRVKELPVLRLDKIAMEKHRGVLECOV---SRA 2316  
2557 ORPKAKPA-----FQSDIA-----FINLT-----EGDTLECKLLITGDP 2590  
2317 SAQVRWFKGSLOFGPKYELV-SDGLYRKLIIISDVHAEDEDVYTC-----DAGDVKTSAQ 2371  
2591 TPFVKWYGTQLVCATEDTEISNANGV-TMKIHGVTADMTGKIKCVAYNKADEVSTEG- 2648  
2372 FFVEEQSITIVRGLQDVTVMEPAWFE---CETSI-----PSVRPPK 2411  
2649 -----PUKVAPIVBFETSUCDRCREGDTLKLRAVLLGEPV---S 2690  
2412 WLLGKTVLQAGNVGLEQEGVHRLMLRTCSNMGVHFV-----GKSRSSARLVVSDI 2467  
2691 WYVNGKLEESONIKIHSEKGYTVTIKIDTCYSGQVCEAINYEGKATSEATL----- 2745  
2468 PVILTRPLEP-----KTG-RELOSVLSCDPRPAPKAVOMYKDDTPLSPSEKPK 2515  
2746 -LVLPREGPPFLWLSNVRARTKVKHVKVFGDPKP---SLTWYINNKELNSDLYT 2801  
2516 MSLEGQMAELRILRMP-----ADAGVYRQCA-----GSAHSSSTVTVVEARE 2557  
2802 IVTDDKTSTLTSNFPDVPVHVEIICAEADAGEVSCANWITYSDMFSESESEAQAE 2861  
2558 -----VTVTGP-----LQDAEATEBGWASFSCLSHDEB---VE 2589  
2862 FVGDDLTEDESLREEMHRTFPMAPKFIITKIDTKAKGHSVAFECV---PDTKGVCCK 2919  
2590 WSLNGMPLYNDSFHEISHKGRHT-----LVLSIORADAG-----IVRASSLKV 2634  
2920 WLKDGKEI-----ELIARIVQTRTGPBGHITQELVLDNVTPEADAGKYTCIVENTAGKD 2973  
2635 STSARLEY-----RVKPVVFKALDLSABERGTLAQCEV-SDPEAHVVVRKDGV- 2684  
2974 TCEATLVIESLEKKSEKKAPEFVALQDTKTKTSKVKLECKVIGEPKPKVSLHDNVS 3033  
2685 -QLGSPDKYDLHTA-----GTRGLVHVDVSPEDAGLYTCHV-----GSEETRAEVRVH 2732  
3034 REKNPSEKITQESITIVESVEGVERVITISSELSHOGKYTCIAENTEGTSKTEAFLTVQ 3093  
2733 DLHVGITRKLTMVLEGECSFECVLSHESASDPAMWTVGGKT-----VGSSSRF--- 2783  
3094 GEAPVFTKELQNKELSIGELVLSCSVKSGPQPHVDYFSPSETTKVETKITSSRIAIEH 3153  
2784 -QATROGKTYI-LVVRBAAPSDAGEVVPVSRGHTSKASLIVREPPAIIKPLEDQWVAPC 2841  
3154 DQTNHWRWISQITKEDIIVSYKAIATNSIGTATSTSKITTKVEAPVFEQGLKKTYSKEK 3213  
2842 EDVRLCSELGRAGTPFWHLKDRKAIRKSKYDVV--CEGTMAMLVIRGASLKDAGEYTCE 2899  
3214 EEIKMEVKVGGSPDVEWFKDDKPVSDGNHMKKNPETGVTFLVVKQAATTDAGKYTAK 3273  
2900 VEASKSTASLHVEKAN-----CFTBEL--TNLQVEEKGTAFTVCTKTEHPAATVTVRKG 2951  
3274 ASNPAGTAESSAAEVVQSLEKPTFVRELVTTEVKINETATLSVTVKGV-PDPSVEWLKD 3332

2952 LLELRASGKHQPSQ-EGJ-TLRLTISALEKADSDTYTC-----DIQQAQSRACLIVQGRV 3005  
3333 GQPVQTDSSHVIKAVEGSGYSITIKDARLEDSGKYACRATNPAGEAKTEANFAVKNLV 3392  
3006 --HIILEDVVDVQGSATFRCRISPANYBPVWFELDKTPLHANELNEIDAQPG-GVHV 3062  
3393 PPEFVEKLSPEVKESESTLISVKVGPPEPSVEWFKDDTTPISIDNHHVIOQTAVGSFS 3452  
3063 LTLRLALKDSOTIYF-----EAGQRAAALRVTEK--PSVFSRELTDAITTEGDELTL 3115  
3453 LTINDARQDVG-IYSCRARNEAGEALTANFGIIRDSIPPEFTQKLRPLEVREOETL 3511  
3116 VCETSTCIP-MCWTWKDKTKLRSARCQLSH-----EGHRAQLLITGALQDSRYK 3167  
3512 KVTVTGTPVNVFVKDDKPI-----NIDNSHIFAKDESGHHT-LTIKQARGEDVGVYTC 3566  
3168 ---EAGACACSSIVRVHAR---PVRFOEALKDLEVEGGAATLRCVLSVAAP-VKWCY 3219  
3567 KATNEAGEAKTANNAVQEEIEAPL-FVOGLKPYEVEGQKPAELVVRVEGKPEPEVKFK 3625  
3220 GNVLRPDGKYSLSROEG--AMLELVRLNRPODSGRYSC-----SFGQOTTSATL----- 3267  
3626 DGVPIADNQHVIEKKGSHGHTLVIKDTNNADPKYTCQATNRAGKDETVGELKIPKYS 3685  
3268 ---TVTALPAQFQIGLKNKEATGATATLRCELSKTA--PVEWRKGSSETLRDGDYCLR 3321  
3686 FEKQTAEBVKPLFTEPLEKETFAVEGDTVVECKVKNKSHPOIKFFKNDQPVIEIQHMLE 3745  
3322 --QDGMCELQIRGLAMVDAABYSC---VGEERTSASLTI----- 3357  
3746 VLEDGNI-KLATQNAKEDVGAYRCEAVNVAGKANTNADLKIOFPAKVVEEHVTDESQLE 3804  
3358 -----RMPBAHFTGLRHQHSIEGATATLRCELSKAAP---VEMRK 3395  
3805 EIQPETVGDTSASTKDTGRAP-EFVELLRSCVTTEKQALLKCV-KGEPRKIKWTK 3862  
3396 -GRESLRDGRHSRQDGAVALCELOICGLAVADAGEYSCVCGEERTSA-----TLTVKAL 3448  
3863 EGKEVMSARVRAEHKDDGTUTLTFDNTVQADAGEYRCEAENEYGSATWTEPIIVTLEGA 3922  
3449 P-----AKFTEGLRNEEAEGATAMLCELS-KVAP-VEMRKGPENLRDGDYCLR--Q 3498  
3923 PKIDGEAPDFLOPKPAVTVVGETAVLEGKISGPKSPVKWYKNGEELKPSDRVKIENLD 3982  
3499 EGTRCELOICGLAMADAGEYLCVCGQE---RTSATLTIRALPAR---FTEDVKNQEAR 3550  
3983 DGTQ-RLTVTNAKLDDMDYRCEASNEFGDVSVDVTLTVKE-PAQVAPGFFKELSAIQVK 4040  
3551 EGATVAVCELANSAAP-VEMRKGSSETLRDGR--YSLRQDQTKCELOIRGLAMADTGEY- 3606  
4041 ETETAFKFCYSGTKPDVKWFKDGTPLKXEDKRHFESTDDGTQ-RLVIEDSKTDQGNR 4099  
3607 ---SCVCCOERTSAMLTVRALP---IKFTEGLRNEEAEGATAVLRCEL-SKMAPVEWVK 3659  
4100 LEVSNDAVANSKVPITV--VPSETLTKKGLTDVNTVQTKILLSVEVEGKPKTVKNYK 4157  
3660 GHETLRDGRHSRQ-DGARCELOIRGLVADAGEYLCMCQKE---RTSAMLTVR----- 3710  
4158 GTETVTSQTTKIVQVTESEYKLEIEGAEAMSDTGAYRVLSTDSFVSSESTATVTVTAAE 4217  
3711 --AMPSKIEGLRNEEAEGDTATLWCEL-SKAPVEMRKGHETLRDGRHSRQDSRC 3767  
4218 KISLPS-FKKGLAQSVKPGTPIVLEIEBKPDKVWYKNGDEIKGKVEDL--GNGKY 4274  
3768 ELQIRGLAVVDAGYSCV---CQOERTSATLTVRALPARFIEDVKNQEAEGATVLOQ 3823  
4275 RLTIIPDFQKDVGEYSVTAANEAGEIESKAKVNVSAKP-EIVSGLVPTTVKQGETATFNV 4333  
3824 EL-SKAAPEVVRKGSSETL-----RGDRYSLRQDQTRCELOIRGLSVADTGEYSCV-- 3873  
4334 KVKGPVKGVKWKNGKEIPDAKTNDGGSYS-----LEIPNAQVEDAADYKVVVS 4384





Db 6038 -----EYNGPEE-----CKPR----- 6049  
QY 5945 YVFRNMKLSIDLNDQVEGDDRAFEVQWQREDSVRKYLQARTAIKSSWVKEICGIIQ 6004  
Db 6050 -IRGLYNS-----IHE----- 6061  
QY 6005 RLALPVRPPDFEBELADCTAELGETVKLACRVGTGPKPVISWKDGKAV---QVDPHHI 6061  
Db 6062 -----GNVEMIVCATGIPTPTVKYKQGEIVGDPGDKRV 6098  
QY 6062 LIEDPDGSCALLDLSLTGVDSPQWCFASAGNCSTLGLV-----QV 6106  
Db 6099 IFTBERGHHUUVNASPDDEGEYSLEATNKLGSAKTEGSLNIIRPHIADABERGMPF 6158  
QY 6107 PPRFVNKVRASGFVEGEDAQCTCTIEGAPQIRKWKDGLATTGNKFTLSBPRSGLLV 6166  
Db 6159 PPGFVRQLKNKHVNHMTIFDCLVVGHPAPEVEMFHNKXIVPGGRIK-IQSCGGSHA 6217  
QY 6167 LVIRAASKEDGLYCECELVNRLGASASABLRIOSPMLQOEQCHREQLVAIVEDTLER 6226  
Db 6218 LIILDTLEDAGEYVATAKNSHGSASSAVLDVTVPFL-----DSIKFN 6261  
QY 6227 ADQEVTSVLKRLG-----PRAPGPTGDLTGP-----PCPRGAP----- 6262  
Db 6262 GEIDVTPLYTEYGFKKLNTASLTPPDPRGPFKEVTGHYLTLSWIPTKAPRYQVTV 6321  
QY 6263 -----ALQETGSGQPPVT-----GTSE-APAVPP----- 6284  
Db 6322 VIRELPEKQWLSLEYNPEFVCKVRNLELGSKYQPRVAENIYIGISDPSAPSRML 6381  
QY 6285 RVQOPLLHE-----GREQEPEATRAQE-----W-----TVPI 6312  
Db 6382 APPQVFDRTNKVIPLLDPVAEKALDNRVSEQACAPWFSQGVVEKRYCAENDTIVL 6441  
QY 6313 RMEG-----AAMPQAGTGELMDVSHVHVRETTORT-----YTYQADTHTA----- 6354  
Db 6442 NVSGFPDPDKWKPRG-----WDI-----DTSPTSCKVYVGGSETTLAITGFSKEN 6490  
QY 6355 -----RPPSMQVIEDVQAOTGGTAQ-----FEALIEGDPQ 6386  
Db 6491 VGQYQCFRANKNDYDQAQNIWDLATRFNIQPLVN-----KTFSSAQPMRMDVRVDGEPP 6546  
QY 6387 SVTKYKDSVLQDSTRLSQQOEGT-TYSLVLRHVASKDAGVYTCLAQNTGGQVLCRAELL 6445  
Db 6547 ELKWKKEWRPIVESSRRIKFDQGPYLCSLIINDPMWRDSDGIYSCVAVNDAGATTCTVT 6606  
QY 6446 VLGDNDPDESKQSHR-----RKLSHYEYKKE 6473  
Db 6607 VEASGDYNDVELPRRVVTIESRRVRELYEISEK 6639

RESULT 5  
T43290  
hemocentin precursor - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 18-Feb-2000  
C:Accession: T43290; T20993; T24734  
R:Vogel, B.E.; Hedgecock, E.M.  
submitted to the EMBL Data Library, June 1998  
A:Description: Hemocentin is required for hemidesmosome mediated cell adhesion and germ-  
A:Reference number: 222396  
A:Accession: T43290  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-5198 <VOG>  
A:Cross-references: EMBL:AF074901; PIDN:AAC26792.1  
R:Sulston, J.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: 219355  
A:Accession: T20993  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5198 <WIL>

A:Cross-references: EMBL:T47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:F15G9.4b  
R:Experimental source: clone F15G9  
R:Kershaw, J.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19929  
A:Accession: T24734  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5198 <W12>  
A:Cross-references: EMBL:T47070; PIDN:CAA87345.1; GSPDB:GN00028; CESP:F15G9.4b  
R:Experimental source: clone T09B9  
C:Genetics:  
A:Gene: him-4; F15G9.4b  
A:Map position: X  
A:Intons: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;  
; 2512/2; 2993/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1;  
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 5100/1  
Query Match 5.0%; Score 2048; DB 2; Length 5198;  
Best Local Similarity 21.4%; Fred. No. 9.8e-53;  
Matches 1187; Conservative 773; Mismatches 2284; Indels 1292; Gaps 254;  
QY 286 VYEDAQ-----ENFVLKILFKOS-----DRG-----LYTCTASNLVQGY-- 321  
Db 136 VFTDARSXDYHLEDEVLTNIQEKSSVVFVMTGCGNTHPGFRTYKIAAASGQVPHL 195  
QY 322 -----SSVLVVVREPA-----VFFKKRLQDLVR-----EKES 349  
Db 196 EKSDYSTVLEYVRRHAVKOKVHLMYEARGGTYSRNPVDKHLSELTISLSDGKDDSDN 255  
QY 350 ATFLCEVQPPQTEAAWFKETRLNASAKYIEEGTERRLTVRNVSDADDAVICETPEG 409  
Db 256 LDIVURDEGRT-----VDRKLYS-----KEGTIDLVKVLIRLKDSPGVWTVNTN 303  
QY 410 SRTVAELAVQGNLARKLPRKTAVRVGDTPMFCVLAVPVPVHWRNQEEVAVAGRVAIS 469  
Db 304 SRLKHTIRVFGHGVDFKYGFASRPLDR-----IELARP-REV-----LNQDVTLLINMTGLI 355  
QY 470 AEGTHTLTISQCCLEDVGQVAFVAGDCQTSRVCVSNAPRKPLOPPVDVVKARMESSV 529  
Db 356 PPGT-----VGEIDLVDYHGHSLYKAVASPHRTNPNMYFAGPPV----- 394  
QY 530 ILSWSPPHGPRPVTIDGYLVEKKLGTYYTWIRCHEAWATPBLTADVAEENGFQFV 589  
Db 395 -----PPKGLFFVRVQGY-----DEDNYEFMR 416  
QY 590 SALNSFGQSPVLEPPGTVHLPKLAVRTPKAVQAVEGVEVTFPSVDITVASAGE--WFLD 647  
Db 417 TAPTAI-----GSVIVGGPRAFMSPIH--QEFVGRDLNLSCTVESASAYTIYVVK 465  
QY 648 GOALKASSVYBIHCDRTHTLTIREVPASLHGA---OLKFVANGIESI--RMEVRAAPG 702  
Db 466 GEDIIGGFLFYHNTDS--VWITIPEL--SLKDAGEYECRVISNNGNYSVKTRVETRESP- 520  
QY 703 LTANKPPAAAAREVLAHHEEAQALLAELSDQAAA---VTWLKDGRTISPGPKYEQVASAG 759  
Db 521 -----PEIFGVNRVNSVPLGEAAFL--HCSTASAGEVBIRWTRYGATVFNTERPNTNG 573  
QY 760 RRVLLVRVDARDDAGLVECVSR--GGRIAYOLSVQGL-----ARFLHDXMAGSCVDVAVAG 813  
Db 574 --TLKIHVTRADAGVTECMARNAGGSTRKRLDIMEPPSVKVTPODVFNMEGV--- 628  
QY 814 PAQFECET-SEAHVHVHWYKDMBLGHSGERFLOEDVGTGRHLVAATVTRQDEGTYSRV 872  
Db 629 --NLSCEAMGDPKPEVHWYFKGRHLLNDYKYVQGD---SKFLYIRDATHHDEGTYECRA 683  
QY 873 -----GEDSVDFRLRVSEPKVVFQKLAARLKQAEAGASATLSCEVAAQ--TEVTWY 924  
Db 684 MSQAGQARDITDML-ATPPKV-----EITQNMVGRGDRVSFECKTIKGGPKIRWF 737  
QY 925 KDGKLSSSSKVCMCATGCTRRLLVVOQAGADAGEYSC----EAGQRLSFHLDV-KEPK 979  
Db 738 KNGKDLIKPDDYIKINEG---QLHIMGAKDEADAGAYSCVGENVAGKDVQVANLSVGRVPT 794









Db 1181 INAEQ---EKIALONDDI-VLECPAKALPPVRLWYEGEKIDSLIPHIREDA--- 1233  
 QY 1405 RRLVQQAQADTGEYSCE---AGQRLSFLSVAB-PKVVFAPKEQVHREVQAQAGAS 1459  
 Db 1234 --LVLQNVKLENTGVFCQVSNLAGEDESLYTLVHEKPKII---SEVPGVVDVVKGFT 1287  
 QY 1460 TTLSCEVAQAQTEYM--WYKDGKLSF--SKVRMEVGCCTRLVVOQAQADAGEYSC-- 1514  
 Db 1288 IEIPCR-ATGVEPVRTWKNKIDILKMEKKFSDNLTUR---IYEAQNDIGNVNV 1343  
 QY 1515 --EAGSQRSLPHVAEPKAVFAKEQEPASREVQAEAGTSATLSCEV-AQAQTEVTVYKDG 1571  
 Db 1344 TNEAGTSQMTTHVDVQEPPIILPSTQNTNTAV---VGRVELKCYVEASPASVTVWFRG 1400  
 QY 1572 KKLSSSK-VRMEVGCCTRLVVOQAQADAGEYSCAGDQ--RLSFLHV---ABPKVV 1625  
 Db 1401 IAGTDTKGYVVEDSG---TLVIOASVEDATIYTCXASNPAKAEANLQVTVIASPDI- 1456  
 QY 1626 FAKQPAHREVQAAGASATLSCEV-AQAQTEVTVYKDGKLSLSSSKVRVEAVGCCTRLV 1684  
 Db 1457 --KOPDVVTQESIKESHPSLYCPVFSNPLQISWYLNKPL-IDDKTSKTSDDKEKLH 1513  
 QY 1695 VQQAQADAGEYSC---EAGQRLSFLHVAELEPOISERPCHREPLVVEHEDILTA 1740  
 Db 1514 VFKAITDSGVYKCVARNAAGEGSKFQVEV--IVPLNLDSEKYYKKVFAKEGEVTLGC 1571  
 QY 1741 TLATPSAATVWLKDGVEIRRSKHEHAS-QGDPTLTVHGAQVLDLSAIVSCRVGAEGQD 1799  
 Db 1572 PVSGFPVQINNVVDGTVVEPKYKGTLSNDGLTLHFDPSVSKQEGNTHCVAQSKGNI 1631  
 QY 1800 FPVQVEVAAKFCRLLEPVCE---LGGVTVLACEL---SPACAEVVMRCGNTQPR 1849  
 Db 1632 LDIDVE---LSVLAVPIVGEDDNLEVLFGKOISLSCDLQTESDDKTTFVWSINGSSED 1686  
 QY 1850 VGKRFQVMAEGPVRSLTVGLRAEDAGEYVCESEDDHTSALQTVSV----- 1895  
 Db 1697 RPDNVQIPSDG--HRLYITDAKPENNGKYMCRVNTSAGKAERTILTLVDLSPFPVPEPVE 1744  
 QY 1896 -----PRVVKFMSG----- 1906  
 Db 1745 ANQKLIKNPILQCVQGNPKPTVIWKIDGNDVDKSWLFDLSLLRIEKLTKSAQIS 1804  
 QY 1907 TVVAEGGEA-----TFQCVSPSDVAVVWRDGA 1936  
 Db 1805 CTAEKAGTASRDFFIQNIAPTFKNEGDOETIFRESEITLDCPVSLGDFQITWKKQGL 1864  
 QY 1937 LLOPSE-----KFAISQSGASHSLTISDLVLEDAGQITVBAEGASSAALRVREAPV 1988  
 Db 1865 PLTENDAIFLTDNLTLILANRDHEDIYTCVANNTAGQVSKDFV-----VQV 1914  
 QY 1989 LFKKKLBPQVVE---ERSSVTELEUTRPWPELWTRNATALAPGNVIEHAGARHL 2044  
 Db 1915 LPKIKNAVTVLEINEGEEIILTCDAE--GNPTAKWDFNQGLD-PKEAVFN---NNHTV 1969  
 QY 2045 VLHNVGADRGFCCGCPDDKTQAKLTVMR---QVRLVRLQAVAREQGTATMEVOLS 2101  
 Db 1970 VVNVTKHTGVYKYATNKVQAVKTVINHVTRKPRFESGLTESELTVNLTSITLECD 2029  
 QY 2102 HADVDG---SWTRDGLRFQOGFTCHLAVRGMHTLT-----LSGLRPEDSG---LMVFK 2149  
 Db 2030 VDDAIGVIGISWVNGKPF-----LAETDGVQTLAGRFLHIVSAKTDHGSYACTVTIN 2082  
 QY 2150 AEGVHTSARLVVTVLPSRFLPDVVTTEKEKVTLECELS-RPNVDVRLKDGVELRAG 2208  
 Db 2083 EAGVATKTNLFVQVPTTVNEGEYTVIENNSLVLPCEVIGRPNPVVWTKDGRPVGDL 2142  
 QY 2209 KTMIAAAGACRSUTIYRCFADQGVVCDADHAQASASVKVQGRVTYTLIRVLAEDAG 2268  
 Db 2143 KSVQVLSG--QQPKIVAEIATKGSYICMAKNDVGTAIESFD---VDIITRPMI----- 2192  
 QY 2269 EIQFVAENAESRAQLRVKELPVTILVR---PLRDKIAMEXHGRVGLSCQVSRASAOVRWF 2324

Db 2193 -----QKGIKNI-VTAIKGALPKFCPIDDDKN-----FKGQIIWL 2228  
 QY 2325 GSO---ELQOPKVELVSDGLYRKLIISDVHAEDEDYTC-----DAGDVTSQAQFFVEE 2376  
 Db 2229 NYQPIDLEAEDARIIRLND---RELTLINTEDEGOYSCRVKNDAGE--NSDFD----- 2279  
 QY 2377 QSITIVRGLODVTVMEPAPAMFECETSIPSRVPPKWLKGTVLQAGGNVGLQEQTVHRL 2436  
 Db 2280 -----KATVLV----- 2286  
 QY 2437 MLRRCTMTGVPVHTFVTKSRSSARLVVSDIPVWLTRLEPKTGRELSVVLSCDRPAP 2496  
 Db 2287 -----PTIIMDKDN-----KTAVEHSTVTLSCPATGK 2316  
 QY 2497 KA-VQVYKDDTFLSPSEKFKMSLEGOM--AELRILRLMPADAGVYRCOAGSAHSTEVTV 2553  
 Db 2317 EBITWFKDGEAIHENTADIIPNGELNGNQLKTRIKEGDAGKTCADNAGSVEODV 2376  
 QY 2554 EAREVTV-----TGPLQDAEATEEGWASFCELS-SHEDEEVEWSLNGMPLYNDSFHEISH 2607  
 Db 2377 NVNVITIPKIEKDGIPSDYESQNERVVISCPVVARPPAKITWLKAGKPLQSDKFKVTS 2436  
 QY 2608 KRRHTLVLSIORADAGIVRASSILKVSARSARLEVRVKPVVPLKALDLSAERSTL--- 2664  
 Db 2437 NGQK--LYLFKLRETDSSKYTCIATNEAGTKDRDFKVSMLV-APSFDPFNIVRITVNSG 2493  
 QY 2665 ---ALQCEV-SDEPAHVVRKDGVLQGPSDKYDFLHTAGTGLVVDVSPEDAGLYTC-- 2718  
 Db 2494 NPSTLHCPAKGSPSTITLWXDGAIEPNDRYVFF-DAG-RQLQISKTEGSDQGRYICIA 2551  
 QY 2719 --HVGSETRARVRVHDHVGITKRLKTMVLEBESCSFECVLSHESASDPAMV---T 2771  
 Db 2552 TNSVSGDLENTLEVIIIPVTDGERREAVAVIEGSELSFC--DSNSTGVDVVEKQDGLT 2609  
 QY 2772 VGGKTVGSSSRFOATROGRKYLIVVREAPSDAGEVWFSVRGLTSKA---SLTVRBRPA 2827  
 Db 2610 INQDILRGDSFIQTPSSGKKMSFL--SARKSDSGRYTCIVRNPAAGEARKLDFDAVNDPFS 2667  
 QY 2828 AI--IKPLEDQWAFGEDVELRCELSSRAGTP-VHMLKDRKAIR-KSQKYDVVCEGTWML 2893  
 Db 2668 ISDELSSANIQIVFPYVPEINCVVSGSPHPKVWLPDDKPLBPDASAAYELTNNGETLKI 2727  
 QY 2884 VIRGASLKDAGEYCEVEAS-----KSTASLHVEEKANCFTBELTNLQVEEXGTA 2933  
 Db 2728 V--RSQVHEAGTYTCEANVGVKARKDFLVRVTPPH-----FEKEREVVARVGDUM 2778  
 QY 2934 VFTCKTEH--PAATVTW-----RKGLLELRASGKHQPSQEGTLRLITISALEKASDST 2984  
 Db 2779 LLTCNAESSVPLSSVYVWHAHDESQNGVI---TSKYAANEK---TLNVNQLDDEGF 2830  
 QY 2985 YTCIDIGQASQAQLLVQGRVHIIEDLEDVDVQE-----GSSATFRCTSPANEPVHW 3038  
 Db 2831 YVC---TAVNEAGITKFKFLIVITETPYLDQOKLYPIILGKRLTLDG--SATCTPPTI 2885  
 QY 3039 FLDKTPLHANELNEIDAQPGYHVHLLRQLAKDSGTIYFE--AGDORASAAALRVTEKPS 3096  
 Db 2886 LFMKDKGRLNSDEVDI--IGSTLVIDNPQKEVEGYTCIAENKAGRSEKONMVEVLLPPK 2944  
 QY 3097 VPSRELTAATTTEBEDTLVCE-TSTCDIPMCWTKD-GKTLRGARQOLSHEGHRAQLLI 3154  
 Db 2945 L-SKEVINVEYQAGDPLTLECPIEDTSGVHTWGRQFKGQQLDMRAQSSD--KSKLYI 3001  
 QY 3155 TGATLQDSGRYKC---EAGGACSSSIVRVHARVVRFOEALKDLEVLGGAATLRCVLSS 3210  
 Db 3002 MQATPEDADSSYCTIAVNDAGAEAVFQVTVNTPKIFGDSFSTTEIVADITLIPCRTEG 3061  
 QY 3211 VAAP-VKWCY-GNNVLR-PGDKYSLRQEGAMLELIVRNLRPDQSGRYSCSFGDQTTSA-- 3265  
 Db 3062 IPPPEISWFLDGKPILEMPGVY---KQGD--LSLRIDNIKFNQEGRYTCVAENKAGRAEQ 3117  
 QY 3266 -TLTVTALPAQFIGKLRKEATEGATATLCEL--SKTAPVWEKSGSETLRDGRY--C 3319  
 Db 3118 DTYVEISEPFRVWASEVVRVVEGRQTIIRCEVGNPEPVWV-----LKGEFYSDL 3171





```

1618 HVAEPKVFABEQ--PAHREVCAGASATISCEVAQAQTEVTWYKDGKLLSS--SSKVR 1673
1400 -----IVDAKNKFIVALKDTEVIEKDDVTLMCTQKTKTGIFNRNKGQISSPPGKFE 1453
1674 VEAVGCTRLVVQAGQADAGEYSCEAGQORLSFLHVAELE--POISERPCR-----RE 1726
1454 TQSRNGTHLTKIGKIEMNEADYIEDQAGLRGSCNVTVLEAKRPIILNWKPKKEAKAGE 1513
1727 PLVVKHEHDIILTALTPSATVTLWLDGVEIIRSEKH--ETASQGDTHTLTVHGQVYL 1784
1514 PCVVKVPQIKGTRR-GDPKQAI-----LKNKGPIDEMRKUVEVIKDDVAEIVFNQOLA 1569
1785 DSAYSXRVG-----AEGQDF-----FVQVVEV 1807
1570 DTGKWAELGNAGTALAPFELFVKDKPKPKGPLETEKXNVTAGGLDLVWGTPDEGAPV 1629
1808 AAKFCRLLE-----PVCELGCTVTLACELSPACAEVVR-----CGNTQPVAGKRFQ 1855
1630 KAYILEMGEGRGNWAKYGETKGTDFVKYKDLKEH--GEYKFRVKALNECGLSDPJTGB--S 1686
1856 MVAEGPVRSLTVGLRAADAGEYVCESDRDHTSAQLTVSVPRVVKFMSGLSTVVAEEGGE 1915
1687 VLAKNP-----YGVPGPKNMDAIDVDKDHCT-----LAWEPPEEDGG 1724
1916 ATPQCVVSPSDVAVVWFRDGLLQPSK-----FAISQSGASHSLTISDLVLEDAQOI-T 1969
1725 API-----TGVIIRERKESEKMDHWQVQTKPDCCCELTDKKVVEDKEYLYR 1769
1970 VEAEGASSAALRVREAPVLFK-KXLEFQ-----TVBERSVTVLELT-RPMPEL 2018
1770 KVAVNKAQPGDCHGKIYKAKKASPEFTGGIKDLRLKVGETIKYDVIFISGEPLPEC 1829
2019 RWTRNATALAPKNVEIHAGARHLVLHNVGFADRGFF-----GCETPDKTQAK 2069
1830 LWWVNGKPLKAVRVKMSSEKGMKIMKTENAVRADSGKFTITLKNSSGSCD-----STAT 1884
2070 LTVENQVRLVRLGLQAVZARQGTATME-----VQLSHADVDSWTRDG 2113
1885 VTVVGRPTPKGFLDIADVCAQD-ATLSNPPDDDDGDLPLTGYIVAEQMDNKNKYIEVG 1943
2114 LRFQOGPTCHLAVRPMHTLISGLRPEDSGLMVFKAEGVHTSARLVVTELPSVFSRPLQ 2173
1944 -----KVDFTITLKVNGLR--NKNYKFRVKAVNEGE-----SEPLS 1980
2174 -DVVTEKEKVTLECELSRP-----NVDVRW--LKQG-----VELRAGKTMAI 2213
1981 ADQYTIKDPWDEPGKPGKPEITDPDADRIDIAEPHPKDGGAPEIBEYIVVRDPDK-- 2038
2214 AAQACRSLTIYRCFEFADQGVVCDADHAQSSASVKVQGRYTLIYRSVLAEADAGEIQFV 2273
2039 -----EWEK-----VKAVPDTNASISGLKEGKEYQPRVAVNKAQFGQ----- 2076
2274 AENASSRAQL-RVKELPVTLRPLRDKIAMEKHGVLCEQVSRASA-----Q 2319
2077 -PSEPSKQLAKPKTIPAWL-----KHDNLKSIIVKAGATVRAWEVKIGGEPIE 2124
2320 VEWKGSQELQPGKVELVSD-----GLYKLIISDVHAEDE----- 2356
2125 VKWFKNGQOLENG--IQITIDTRKNEHTILCIPSMRSDVGEYR-LTVKNSHGADEKAK 2181
2357 ----DITYTCDAGDVXTSAQFFVEQSIITVIRGLQDVTYMEP-----APAWPEC 2400
2182 LTVLDRPSKPNGLSVSDVF--EDNLNLSMKPPDDDGGEPIEYVEVEKLOTATGRWVPC 2238
2401 -----ETSIPIV-- 2407
2239 AKVKDTKAHIDGLKXGQTYQFRVKAVNKEGASDALSTDKDTKAKNPYDEPGKTGTDPDVD 2298
2408 -----RPPKWLGLKTVLQ-----AGNVUGLE-----QEGTVH 2434
2299 WDADRVSLEWEPKSDGGAPITQYVIEKKXGHRDWOECGVKSGDQTNABILGLKEGEY 2358
2435 RLMLRRTCTMTGPHFTVTKSRSSARLVVSDIPVVLTRPLEPKTKRELQSVVL-----S 2489
2359 QPRVKAV--NKAGP-----GEASDPKRVKAV-----PRLKFWIDREAMKITIKVGN 2406
2490 CDF-----RPAPKAVQVYKDDTPLSPSEKFKMSLEGOAELRILMLPADAGVRCQA- 2542
2407 VEPDVPVRGPPPKK-EWIFNEKFPD-DQKIRIESDYKTRFVLRGATKHAAGLYTLTAT 2464
2543 ---GSAHSSTEVTEAREVTVTGPLODAEATEGMAWFSCELSHEDDEBEWESL-----NGM 2595
2465 NASGSKHSEVIVLJKPSPPLGLEVSNVYE-----DRADLEWKVPEDDGA 2512
2596 PLYNDSFHEISHK-----GRHHTLVLSIQRADAGIVRASSLSKYSTSARLEVRVK 2645
2513 PI---DHYETEKMDLATGRWVPCGRSET-----TKTTVPNLQPGHEYKFRVR 2556
2646 PVVFLKALDLSAE---ERGTALACEVSDPEAHVVWRKDGVLQGL----- 2687
2557 AVNKEGSDPLTNTNTAILAKNPYEPGKVDKPEL-VWDKDHVDLAWNAPDDGGAGIEAF 2615
2688 -----PSD-----KYDFLHTAGTRGLVHVHDSVPEDAGLYT 2717
2616 VIEKKDKNGRWEALVVPDQKTATVPNLKEGEYQFRISARNKAGTGDPSDPSD----- 2670
2718 CHVGSETRARVRVHDLHVGIITKRLKTMVELEGESCFEVLVSHESASDPAMKTVCGKIV 2777
2671 -RVAKPRNLAPRIH-----REDLSDTTVKVGATLKFIVHIDGEPAPD-VTWSFNGKI 2722
2778 GSS-----SRF-----QATROGRKYLVRREAPSDAGEVWFSVRGLTSK----- 2817
2723 GESKAQIENEPYISRALPKALRKQSGKTYIATNNGTDSVTINIKVSKPTKPKGPIE 2782
2818 ASLIVREPAAIKPLEDQ-----WV----- 2838
2783 VTDVFEDRATLWKPPEDDGGEPIDFYEIEKMNKDGIVWPCGRSGDTHFTVDSLKNKG 2842
2839 -----APGEDVELRCEL-----SRAGTP-----VHM--LKDRK 2864
2843 YKFRVKAIVNSGSDPLETETDILAKNPDRDPRDPRPEPTWDSHVLDLKWDPPLSDGG 2302
2865 A-----IRKSQYD-----VVCCEGTAMLVIRGASLKDAGEYTCV-----E 2901
2903 AP-EEYOIEKRTKYGRWEPAITVPGQTTATV--PDLTPNEEYEFVRVAVNKGSPDPS 2960
2902 ASKSTASLHVEEKANCFTELTNLQVEEKGTAFTCKTE-HPAATVTWR-KGULELRAG 2959
2961 ASKAVTAKPRNLKPHIDRALKNLTIKAGCSISFVFPVSGEPAPTVTWHWPNRREIRNG 3020
2960 KHQPSQEGTLRLTISALEKADSDTYTCDIGQAQSRALLVOGRRVHIIIE-----D 3010
3021 RVKLDNPEYQSKLVVQKMERGDSGTFTIKAVNANGDEATV--KINVIDKPTSPNGPLD 3077
3011 LEDVD-----VOEGSSAT----- 3023
3078 VSDVHGDHVTLANWRAPDDGGIPIENYVIEKYDTASGRWVPAKAVAGDKTTAVVDGLIFG 3137
3024 --PRCISIPANYB---PVHWF---LDKTLPL-----HANELNEID-----AQP 3057
3138 HEYKFRVAVNAEGESDPLETGTCTLLAKDPFPKPKGTNAPEITDWDKDHVDLEWKPPAND 3197
3058 GG-----YHVLTLRQLA-----LKDSGTIYF-----EAG-----D 3082
3198 GGAPIEYVEMKDEFPWNDVAHPAGQTNATVGNLXEGSKYEPRIRAKNKAAGLDPS 3257
3083 QRASAAALRVTEKPSVFSR-ELTDATITEGEDTLVCETSTCDIP-MCWTIKDGKTLRGSAR 3140
3258 DSASAVAKARNVPVDIRNSIQEIKVQAGDPSLNPVSGEPTPTITWTFEPTVPSDDR 3317
3141 COLSHGHRQAQLITGATLQDSGRY--KCEAGACSSSIVRVAH-----RPVRFOEA 3190
3318 MCLNNEGDKTKFVKRALRSDTGTYYIIKAENENGTDTAEVKVTVLDPHPSPRGLDVTNI 3377
3191 LKDLVL-----EGGAATLRCVLSSV-AAPVKW-----CVGNV----- 3222

```



Db 3378 VKDGCDAWKPEDDGGAEISHYVIERQDAATGWTAGGESKDNFHVDDLTQGEYKPR 3437  
 QY 3223 ---VLRPOCKYSLRQEGAML---ELV---VRNLRPODSGRYSCS 3257  
 Db 3438 KVAVNRHGSDPLREAREAAIAKDPFDRADKPGTPEIVDWKDHADLKWTPPADGGAPIE 3497  
 QY 3258 FGDQTTSATLTVPALPAQFICKLRNKGATEGATLRLCELSKTAPVWRKGSSETLRDGR 3317  
 Db 3498 -GYLVEMRTPSGDWVPAVTVG---AGE-LTATV---DGLKPGQT 3533  
 QY 3318 YCLRQDGMCELOIRGLAMVDAAYSCVCGEERTSASLTIRPMPAHFIBLRH---OE 3372  
 Db 3534 Y-----QPRVKNALNAGE-----STPDSRRTMVAKPRHLAPKINDFMFAQR 3576  
 QY 3373 SIEGATAT--LRCELSKAAPVEM-RKGRESLRDGRSLRQDGAVCBLQICGLAVADAGE 3429  
 Db 3577 VKAGOTLNFVNVGEPAKPIEWFLNGSPSSGGNTHIDNNTDNTKLTXTARADSGK 3636  
 QY 3430 YSCVC---GBERTSATLTVKALPAKFTFEGURNBEAVEGATAMLWCELSKVAPVWRKGP 3485  
 Db 3637 YKIVATNESGKDEHEVDNILDIPGAPEGLRHKDITKESVULKWDE---PLDDGGSP 3691  
 QY 3486 ---ENLRDGRVILLRQETRCLOELQICGLAMADAGEY---LCVCGEERTSATLTI-RA 3535  
 Db 3692 ITNVVVEKQEDGGRWVPCGETSDTSLKVNKL--SEGHEYKFRVAVNRQGTSAPLTSDHA 3749  
 QY 3536 LPARFIEDVKQAEAREGATVLOCELSAAPVEMVRKGSSETL---RDG---DRYSL- 3584  
 Db 3750 IVA---KNPDEPDAP---TDVTPVDMDKDHVLEWKPXPANDGGAPIDAYIVE 3796  
 QY 3585 ---RDG--TKC---ELQIRGLAMADTGEYS---CYCQER 3614  
 Db 3797 KDKFGDWVECARVDGKTTKATADNLNTPGETYQFRVAVNRKAGPKSPDPTGNVAKPRR 3856  
 QY 3615 TSAMLTVRA---LPIKFTFEGURNEBEATEGATVRLCELSKMAPVEMWKGEH-TLRDGRH 3670  
 Db 3857 MAPKLNLAGLLDLRAIKAGTPKIDIAPEG-----EPAPVAKMKANDATIDTGARA 3906  
 QY 3671 SLRQDGAECLEIQRGLVAEDAGEVLCMC---GKERTSAMLTVRAMPKSTFIEGLURNEEA- 3725  
 Db 3907 DVTNPTSSAIIHFSAVRGDTGVYKIIIVENEHGHKDTAQCNVTVLDVFGTPEGLKIDEIH 3966  
 QY 3726 TEGDTATLWCELSKAAAPVEMWKGHETLR-----DGRHSLRQDGSRCLEIQRGLAVVDA 3779  
 Db 3967 KEG-----CTLNWKPFDD-NGGTDVLYHIVKMDTSRGTWQEVGTFPDCTAKNKLVP 4019  
 QY 3780 GEYSCVCGEERTSA-TLTVRALPARFIE-----DVK-N 3810  
 Db 4020 KEYAF-----RVKAVNLGSGKPLEAEPIIAKNQFQVDPDPKPEVTDMDKORIDIKWN 4074  
 QY 3811 QEARREGATVLIQ--CELSKAAPVEMVRKGSSETLRGDRYSL--RODGTRECELOIHLGSVAD 3966  
 Db 4075 PTANNGGAPVTGYIVEKKEKESAIWTERAGKT--PGTTFSDNLKPGVEYEFVRVAVNAAG 4132  
 QY 3867 TGEYSCVCGEERTSATLTVPAPQVFPREPLQSLQAEGSGSTATLQCE-LSEPTATVVMWSK 3925  
 Db 4133 PSDPSDPTDPTTKA---RYLKPKILITASRKIKIKAGFTHNLEVDVFIGAPDPTATWTVG 4188  
 QY 3926 GLQIQANGREPRRL-----QGTABLVLDLOREDTGBYTCCTGSGQATSATLTVTAAVRF 3981  
 Db 4189 ---DSGAALAPPELLVDAKSTTSIFFPSAKRADSGNY-----KLKV 4226  
 QY 3982 LRELQHQZ-----VDEGGTAHLCELSRAGASVEMVRKGSILQLPPCAKQWV-----QD 4029  
 Db 4227 KNELGEDEAIFEVIVQDRPSA-----PEGPLEVSDVTKDSCVLNWKPPKD 4271  
 QY 4030 GAABELLVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFTRLQSLQEBQETGDIARLCC 4089  
 Db 4272 DGGAEISNVVVEKBDTKNTW-----VPVSAPVT----- 4300  
 QY 4090 QLSDAESGAVVQWLKEGVHELHAGPKYEMRSQCATRELLIHOLEAKDTEYACVTTGGOKTA 4149

Db 4301 ----GTSITVPKLTGEGHE-----YEFR---VVAENTFGRSDSLNTDEPVLAKDPFGTP 4346  
 QY 4150 ASLRVTEPEVTVIRGLVDAEVTADDEDVEFS-----CEVRAGATGVQWC-L 4194  
 Db 4347 G--KGRPEI-----VD---TDNDHIDLKWDPPRDNGSGSPVDHYDIERKDAKTRWIKV 4395  
 QY 4195 QGLPIQSNNEVTEVAVRDGRIHTLRUKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVITLE 4254  
 Db 4396 NTSVPQGTAFSTRVQKG--HTYEYRVAVNKAQPGQPSDSSAAATAKAPMEAPKFDL 4453  
 QY 4255 PLQDVQLSEGQDASQCRLSRASQOEABWAL--GGVPLQANEMNDITVQGTLLHLLTHKV 4313  
 Db 4454 DGKEFRVAKAGEPLVITITFTASQPQDISWTEGGRPLAGVETTDSQTK-----LVIPST 4507  
 QY 4314 TLEDAGTVSFHVTG--CSSEAOIKVTAKNTVVRGLENYEALLEGGBALFECOLSQPEVAH 4371  
 Db 4508 RRSDSPVKIKAVNPFYGEAEANIKIV-----IDKGAPE-----NITYPAVSRH 4552  
 QY 4372 TWLLDDEPVRITSENAEVVFFENGLEHLLHLLKMLRPQDSCTVTFIAGDMVTSATLTVGRW 4431  
 Db 4553 TCTLNWDAPKDDGGAEIAGYKIEYQEV-----GSQWDKVPGLISGTAYTVRG-- 4600  
 QY 4432 LBILEPLKNAAVRAGAQAQARFTCTLSEAVPVGGEASWYINGAAV--QPDDSDMTVTADGSH 4488  
 Db 4601 -----LEHCQQXRFRIAEANAVGLSD---YCOGVFWIKDPDP----- 4636  
 QY 4489 QALLRSAPQHAGVETFACRDVAVASARLTVLGLPDPPEDEAVVAHSSHTVTLWSAAPMS 4548  
 Db 4637 -----PGAPSTPEITGYDTNQVSLAMNPPRD 4662  
 QY 4549 DGGGGLGCVRYVE--VKEGATGQWR--LCHELVPGPEVVDGLAPGETYFRVAAVGPVCG 4605  
 Db 4663 DGSFILGVVBERFERGG--GDWAPVPMVMWKGTECIVPGLHEHETQYFRVAVNAAGH 4721  
 QY 4606 EPVHLPTVRLAEPKPPVPPQSPAPESPQVAAGEDVSLELVAVAEAGVIVHKGWRIQ 4665  
 Db 4722 EPSNGSEPTC---RPVYKPGAPDAPRG-----KITKNSABELTWNR---PLRD 4765  
 QY 4666 GG-----RFEVVSQGRQOMLVIKGFTAEDQGEYHCG 4696  
 Db 4766 GAPIDIGYIVEKKLGDNMTKNDKVPDRDTAFEVKNLGEKE-----EYEF 4812  
 QY 4697 LAQGSICPAATFQVALSPASVDEAPQPSLPPEAAQEGDLHLLWEALARKERMSREPTLD 4756  
 Db 4813 V-----JAVNSAGEGEPKPS-----DLVLEBQPCR-----PIFD 4843  
 QY 4757 --SISELPEEDORS--QRLPQEAEEVAP--DLSEGYSTADELART-----GDADLSHTS 4804  
 Db 4844 INNLKDDITVRAGETTQIRIPYAGGNPKPIIDLFGNSPFIENERTVVDVNPGEIVITTG 4903  
 QY 4805 SDESAGTSPSLVTLKXAGR-----PG--TSPL-ASKVGAPAAP-SYKPOQ 4848  
 Db 4904 S-KRSDAG-PYKISATNKYKDTCKLVNFVLDAPGKPTGPIRATDIQADAMTSLWRPCKD 4961  
 QY 4849 Q-----EPLAAVRPLDGLSTKDLGDFSMKAAVKIOAAPKGVKVRKEMKQOQBPESH 4902  
 Db 4962 NGGDATITVYWEKRTFGGDWVT--VGHEVGTTLVRNLDANTPTYFRVRAENQYV----- 5014  
 QY 4903 TFGDTEAOVGDALRL-----ECVVASKADVRARWLK---DG----- 4935  
 Db 5015 -----VGEFLETTDAIVAKNPPDTFGAPQGPFAVETSEATLQWTRPDSGGAPIQ 5066  
 QY 4936 ---VE-----LTDGRHHHIDQLDGTCSLLIAGLDRADAGCYTCQVSNK 4976  
 Db 5067 GVVIKREYGVSTEWTKAAAFGNILDTKRVGTGLTKPKTYEFVRVAAVNAAGQGEYS----- 5120  
 QY 4977 FQGVTHSACVVVSGSESESSGCGELDDAFRRARRLHRLFR-TKSPAESVSDDELFLSA 5035  
 Db 5121 ----VNSVPTADNAPTFRKINMGMLTRDILAYAGERAKILVPPAASPA----PKVTFSK 5172  
 QY 5036 DEGPAPPEPADWQTYREDEHFICIRFBALEARQAVTRFOEMFATIGIVEIKLVEQ-- 5093  
 Db 5173 GENKISPTDPRVKVEYSDFLATLTIEKSELTDGGLYFVELENSQGSDSASIRLKVWDKPA 5232





A:Accession: S07571  
 A:Molecule type: DNA  
 A:Residues: 792-6839 <BEN2>  
 A:Cross-references: EMBL:X15423; NID:g6897; PIDN:CAA33463.1; PID:g6898  
 A:Experimental source: var. Bristol  
 R:Benian, G.M.; Kiff, J.B.; Neckelmann, N.; Moerman, D.G.; Waterston, R.H.  
 Nature 342, 45-50, 1989  
 A:Title: Sequence of an unusually large protein implicated in regulation of myosin activity  
 A:Reference number: S06797; MUID:90044042; PMID:2812002  
 A:Accession: S06797  
 A:Molecule type: DNA  
 A:Status: nucleic acid sequence not shown  
 A:Residues: 806-1175,1178-1998,'Y',2000-3040,'I',3042-3335,'I',3337-5693,5696-6359,'I',6361-6839  
 A:Cross-references: EMBL:X15423  
 A:Experimental source: var. Bristol  
 R:Benian, G.M.; L'Hernault, S.W.; Morris, M.E.  
 Genetics 134, 1097-1104, 1993  
 A:Title: Additional sequence complexity in the muscle gene, unc-22, and its encoded protein  
 A:Reference number: S57218; MUID:93387664; PMID:8397135  
 A:Accession: S57218  
 A:Molecule type: DNA  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Residues: 2-99,108-194,'Q',196-206,374-468,658-753 <BEN4>  
 A:Experimental source: var. Bristol  
 R:White, S.  
 submitted to the EMBL Data Library, May 1996  
 A:Reference number: Z20442  
 A:Accession: Z27934  
 A:Molecule type: DNA  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Residues: 'MGIPGKCKQ',19-6839 <WIL>  
 A:Cross-references: EMBL:Z73897; PIDN:CAA98064.1; GSPDB:GN00022; CESP:ZK617.1a  
 A:Experimental source: clone ZK617  
 R:Harris, B.  
 submitted to the EMBL Data Library, May 1996  
 A:Reference number: Z20458  
 A:Accession: Z28030  
 A:Molecule type: DNA  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Residues: 'MGIPGKCKQ',19-6839 <WIL>  
 A:Cross-references: EMBL:Z73899; PIDN:CAA98081.1; GSPDB:GN00022; CESP:ZK617.1a  
 A:Experimental source: clone ZK829  
 C:Comment: Lack of unc-22 leads to a constant twitching of the body muscles.  
 C:Genetics:  
 A:Gene: unc-22; CESP:ZK617.1a  
 A:Map position: 4  
 A:Introns: 18/3; 69/3; 143/2; 176/3; 264/2; 387/3; 413/2; 471/1; 516/3; 550/3; 582/3; 601/3; 669/3; 676/3; 680/3  
 C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;  
 C:Keywords: ATP; autophosphorylation; duplication; muscle; phosphotransferase; serine/threonine kinase  
 F:806-898,899-990,991-1083,1084-1175,1178-1273,1474-1567,1770-1864,2066-2158,2358-2450,2451-2452,2453-2454,2455-2456,2457-2458,2459-2460,2461-2462,2463-2464,2465-2466,2467-2468,2469-2470,2471-2472,2473-2474,2475-2476,2477-2478,2479-2480,2481-2482,2483-2484,2485-2486,2487-2488,2489-2490,2491-2492,2493-2494,2495-2496,2497-2498,2499-2500,2501-2502,2503-2504,2505-2506,2507-2508,2509-2510,2511-2512,2513-2514,2515-2516,2517-2518,2519-2520,2521-2522,2523-2524,2525-2526,2527-2528,2529-2530,2531-2532,2533-2534,2535-2536,2537-2538,2539-2540,2541-2542,2543-2544,2545-2546,2547-2548,2549-2550,2551-2552,2553-2554,2555-2556,2557-2558,2559-2560,2561-2562,2563-2564,2565-2566,2567-2568,2569-2570,2571-2572,2573-2574,2575-2576,2577-2578,2579-2580,2581-2582,2583-2584,2585-2586,2587-2588,2589-2590,2591-2592,2593-2594,2595-2596,2597-2598,2599-2600,2601-2602,2603-2604,2605-2606,2607-2608,2609-2610,2611-2612,2613-2614,2615-2616,2617-2618,2619-2620,2621-2622,2623-2624,2625-2626,2627-2628,2629-2630,2631-2632,2633-2634,2635-2636,2637-2638,2639-2640,2641-2642,2643-2644,2645-2646,2647-2648,2649-2650,2651-2652,2653-2654,2655-2656,2657-2658,2659-2660,2661-2662,2663-2664,2665-2666,2667-2668,2669-2670,2671-2672,2673-2674,2675-2676,2677-2678,2679-2680,2681-2682,2683-2684,2685-2686,2687-2688,2689-2690,2691-2692,2693-2694,2695-2696,2697-2698,2699-2700,2701-2702,2703-2704,2705-2706,2707-2708,2709-2710,2711-2712,2713-2714,2715-2716,2717-2718,2719-2720,2721-2722,2723-2724,2725-2726,2727-2728,2729-2730,2731-2732,2733-2734,2735-2736,2737-2738,2739-2740,2741-2742,2743-2744,2745-2746,2747-2748,2749-2750,2751-2752,2753-2754,2755-2756,2757-2758,2759-2760,2761-2762,2763-2764,2765-2766,2767-2768,2769-2770,2771-2772,2773-2774,2775-2776,2777-2778,2779-2780,2781-2782,2783-2784,2785-2786,2787-2788,2789-2790,2791-2792,2793-2794,2795-2796,2797-2798,2799-2800,2801-2802,2803-2804,2805-2806,2807-2808,2809-2810,2811-2812,2813-2814,2815-2816,2817-2818,2819-2820,2821-2822,2823-2824,2825-2826,2827-2828,2829-2830,2831-2832,2833-2834,2835-2836,2837-2838,2839-2840,2841-2842,2843-2844,2845-2846,2847-2848,2849-2850,2851-2852,2853-2854,2855-2856,2857-2858,2859-2860,2861-2862,2863-2864,2865-2866,2867-2868,2869-2870,2871-2872,2873-2874,2875-2876,2877-2878,2879-2880,2881-2882,2883-2884,2885-2886,2887-2888,2889-2890,2891-2892,2893-2894,2895-2896,2897-2898,2899-2900,2901-2902,2903-2904,2905-2906,2907-2908,2909-2910,2911-2912,2913-2914,2915-2916,2917-2918,2919-2920,2921-2922,2923-2924,2925-2926,2927-2928,2929-2930,2931-2932,2933-2934,2935-2936,2937-2938,2939-2940,2941-2942,2943-2944,2945-2946,2947-2948,2949-2950,2951-2952,2953-2954,2955-2956,2957-2958,2959-2960,2961-2962,2963-2964,2965-2966,2967-2968,2969-2970,2971-2972,2973-2974,2975-2976,2977-2978,2979-2980,2981-2982,2983-2984,2985-2986,2987-2988,2989-2990,2991-2992,2993-2994,2995-2996,2997-2998,2999-3000,3001-3002,3003-3004,3005-3006,3007-3008,3009-3010,3011-3012,3013-3014,3015-3016,3017-3018,3019-3020,3021-3022,3023-3024,3025-3026,3027-3028,3029-3030,3031-3032,3033-3034,3035-3036,3037-3038,3039-3040,3041-3042,3043-3044,3045-3046,3047-3048,3049-3050,3051-3052,3053-3054,3055-3056,3057-3058,3059-3060,3061-3062,3063-3064,3065-3066,3067-3068,3069-3070,3071-3072,3073-3074,3075-3076,3077-3078,3079-3080,3081-3082,3083-3084,3085-3086,3087-3088,3089-3090,3091-3092,3093-3094,3095-3096,3097-3098,3099-3100,3101-3102,3103-3104,3105-3106,3107-3108,3109-3110,3111-3112,3113-3114,3115-3116,3117-3118,3119-3120,3121-3122,3123-3124,3125-3126,3127-3128,3129-3130,3131-3132,3133-3134,3135-3136,3137-3138,3139-3140,3141-3142,3143-3144,3145-3146,3147-3148,3149-3150,3151-3152,3153-3154,3155-3156,3157-3158,3159-3160,3161-3162,3163-3164,3165-3166,3167-3168,3169-3170,3171-3172,3173-3174,3175-3176,3177-3178,3179-3180,3181-3182,3183-3184,3185-3186,3187-3188,3189-3190,3191-3192,3193-3194,3195-3196,3197-3198,3199-3200,3201-3202,3203-3204,3205-3206,3207-3208,3209-3210,3211-3212,3213-3214,3215-3216,3217-3218,3219-3220,3221-3222,3223-3224,3225-3226,3227-3228,3229-3230,3231-3232,3233-3234,3235-3236,3237-3238,3239-3240,3241-3242,3243-3244,3245-3246,3247-3248,3249-3250,3251-3252,3253-3254,3255-3256,3257-3258,3259-3260,3261-3262,3263-3264,3265-3266,3267-3268,3269-3270,3271-3272,3273-3274,3275-3276,3277-3278,3279-3280,3281-3282,3283-3284,3285-3286,3287-3288,3289-3290,3291-3292,3293-3294,3295-3296,3297-3298,3299-3300,3301-3302,3303-3304,3305-3306,3307-3308,3309-3310,3311-3312,3313-3314,3315-3316,3317-3318,3319-3320,3321-3322,3323-3324,3325-3326,3327-3328,3329-3330,3331-3332,3333-3334,3335-3336,3337-3338,3339-3340,3341-3342,3343-3344,3345-3346,3347-3348,3349-3350,3351-3352,3353-3354,3355-3356,3357-3358,3359-3360,3361-3362,3363-3364,3365-3366,3367-3368,3369-3370,3371-3372,3373-3374,3375-3376,3377-3378,3379-3380,3381-3382,3383-3384,3385-3386,3387-3388,3389-3390,3391-3392,3393-3394,3395-3396,3397-3398,3399-3400,3401-3402,3403-3404,3405-3406,3407-3408,3409-3410,3411-3412,3413-3414,3415-3416,3417-3418,3419-3420,3421-3422,3423-3424,3425-3426,3427-3428,3429-3430,3431-3432,3433-3434,3435-3436,3437-3438,3439-3440,3441-3442,3443-3444,3445-3446,3447-3448,3449-3450,3451-3452,3453-3454,3455-3456,3457-3458,3459-3460,3461-3462,3463-3464,3465-3466,3467-3468,3469-3470,3471-3472,3473-3474,3475-3476,3477-3478,3479-3480,3481-3482,3483-3484,3485-3486,3487-3488,3489-3490,3491-3492,3493-3494,3495-3496,3497-3498,3499-3500,3501-3502,3503-3504,3505-3506,3507-3508,3509-3510,3511-3512,3513-3514,3515-3516,3517-3518,3519-3520,3521-3522,3523-3524,3525-3526,3527-3528,3529-3530,3531-3532,3533-3534,3535-3536,3537-3538,3539-3540,3541-3542,3543-3544,3545-3546,3547-3548,3549-3550,3551-3552,3553-3554,3555-3556,3557-3558,3559-3560,3561-3562,3563-3564,3565-3566,3567-3568,3569-3570,3571-3572,3573-3574,3575-3576,3577-3578,3579-3580,3581-3582,3583-3584,3585-3586,3587-3588,3589-3590,3591-3592,3593-3594,3595-3596,3597-3598,3599-3600,3601-3602,3603-3604,3605-3606,3607-3608,3609-3610,3611-3612,3613-3614,3615-3616,3617-3618,3619-3620,3621-3622,3623-3624,3625-3626,3627-3628,3629-3630,3631-3632,3633-3634,3635-3636,3637-3638,3639-3640,3641-3642,3643-3644,3645-3646,3647-3648,3649-3650,3651-3652,3653-3654,3655-3656,3657-3658,3659-3660,3661-3662,3663-3664,3665-3666,3667-3668,3669-3670,3671-3672,3673-3674,3675-3676,3677-3678,3679-3680,3681-3682,3683-3684,3685-3686,3687-3688,3689-3690,3691-3692,3693-3694,3695-3696,3697-3698,3699-3700,3701-3702,3703-3704,3705-3706,3707-3708,3709-3710,3711-3712,3713-3714,3715-3716,3717-3718,3719-3720,3721-3722,3723-3724,3725-3726,3727-3728,3729-3730,3731-3732,3733-3734,3735-3736,3737-3738,3739-3740,3741-3742,3743-3744,3745-3746,3747-3748,3749-3750,3751-3752,3753-3754,3755-3756,3757-3758,3759-3760,3761-3762,3763-3764,3765-3766,3767-3768,3769-3770,3771-3772,3773-3774,3775-3776,3777-3778,3779-3780,3781-3782,3783-3784,3785-3786,3787-3788,3789-3790,3791-3792,3793-3794,3795-3796,3797-3798,3799-3800,3801-3802,3803-3804,3805-3806,3807-3808,3809-3810,3811-3812,3813-3814,3815-3816,3817-3818,3819-3820,3821-3822,3823-3824,3825-3826,3827-3828,3829-3830,3831-3832,3833-3834,3835-3836,3837-3838,3839-3840,3841-3842,3843-3844,3845-3846,3847-3848,3849-3850,3851-3852,3853-3854,3855-3856,3857-3858,3859-3860,3861-3862,3863-3864,3865-3866,3867-3868,3869-3870,3871-3872,3873-3874,3875-3876,3877-3878,3879-3880,3881-3882,3883-3884,3885-3886,3887-3888,3889-3890,3891-3892,3893-3894,3895-3896,3897-3898,3899-3900,3901-3902,3903-3904,3905-3906,3907-3908,3909-3910,3911-3912,3913-3914,3915-3916,3917-3918,3919-3920,3921-3922,3923-3924,3925-3926,3927-3928,3929-3930,3931-3932,3933-3934,3935-3936,3937-3938,3939-3940,3941-3942,3943-3944,3945-3946,3947-3948,3949-3950,3951-3952,3953-3954,3955-3956,3957-3958,3959-3960,3961-3962,3963-3964,3965-3966,3967-3968,3969-3970,3971-3972,3973-3974,3975-3976,3977-3978,3979-3980,3981-3982,3983-3984,3985-3986,3987-3988,3989-3990,3991-3992,3993-3994,3995-3996,3997-3998,3999-4000,4001-4002,4003-4004,4005-4006,4007-4008,4009-4010,4011-4012,4013-4014,4015-4016,4017-4018,4019-4020,4021-4022,4023-4024,4025-4026,4027-4028,4029-4030,4031-4032,4033-4034,4035-4036,4037-4038,4039-4040,4041-4042,4043-4044,4045-4046,4047-4048,4049-4050,4051-4052,4053-4054,4055-4056,4057-4058,4059-4060,4061-4062,4063-4064,4065-4066,4067-4068,4069-4070,4071-4072,4073-4074,4075-4076,4077-4078,4079-4080,4081-4082,4083-4084,4085-4086,4087-4088,4089-4090,4091-4092,4093-4094,4095-4096,4097-4098,4099-4100,4101-4102,4103-4104,4105-4106,4107-4108,4109-4110,4111-4112,4113-4114,4115-4116,4117-4118,4119-4120,4121-4122,4123-4124,4125-4126,4127-4128,4129-4130,4131-4132,4133-4134,4135-4136,4137-4138,4139-4140,4141-4142,4143-4144,4145-4146,4147-4148,4149-4150,4151-4152,4153-4154,4155-4156,4157-4158,4159-4160,4161-4162,4163-4164,4165-4166,4167-4168,4169-4170,4171-4172,4173-4174,4175-4176,4177-4178,4179-4180,4181-4182,4183-4184,4185-4186,4187-4188,4189-4190,4191-4192,4193-4194,4195-4196,4197-4198,4199-4200,4201-4202,4203-4204,4205-4206,4207-4208,4209-4210,4211-4212,4213-4214,4215-4216,4217-4218,4219-4220,4221-4222,4223-4224,4225-4226,4227-4228,4229-4230,4231-4232,4233-4234,4235-4236,4237-4238,4239-4240,4241-4242,4243-4244,4245-4246,4247-4248,4249-4250,4251-4252,4253-4254,4255-4256,4257-4258,4259-4260,4261-4262,4263-4264,4265-4266,4267-4268,4269-4270,4271-4272,4273-4274,4275-4276,4277-4278,4279-4280,4281-4282,4283-4284,4285-4286,4287-4288,4289-4290,4291-4292,4293-4294,4295-4296,4297-4298,4299-4300,4301-4302,4303-4304,4305-4306,4307-4308,4309-4310,4311-4312,4313-4314,4315-4316,4317-4318,4319-4320,4321-4322,4323-4324,4325-4326,4327-4328,4329-4330,4331-4332,4333-4334,4335-4336,4337-4338,4339-4340,4341-4342,4343-4344,4345-4346,4347-4348,4349-4350,4351-4352,4353-4354,4355-4356,4357-4358,4359-4360,4361-4362,4363-4364,4365-4366,4367-4368,4369-4370,4371-4372,4373-4374,4375-4376,4377-4378,4379-4380,4381-4382,4383-4384,4385-4386,4387-4388,4389-4390,4391-4392,4393-4394,4395-4396,4397-4398,4399-4400,4401-4402,4403-4404,4405-4406,4407-4408,4409-4410,4411-4412,4413-4414,4415-4416,4417-4418,4419-4420,4421-4422,4423-4424,4425-4426,4427-4428,4429-4430,4431-4432,4433-4434,4435-4436,4437-4438,4439-4440,4441-4442,4443-4444,4445-4446,4447-4448,4449-4450,4451-4452,4453-4454,4455-4456,4457-4458,4459-4460,4461-4462,4463-4464,4465-4466,4467-4468,4469-4470,4471-4472,4473-4474,4475-4476,4477-4478,4479-4480,4481-4482,4483-4484,4485-4486,4487-4488,4489-4490,4491-4492,4493-4494,4495-4496,4497-4498,4499-4500,4501-4502,4503-4504,4505-4506,4507-4508,4509-4510,4511-4512,4513-4514,4515-4516,4517-4518,4519-4520,4521-4522,4523-4524,4525-4526,4527-4528,4529-4530,4531-4532,4533-4534,4535-4536,4537-4538,4539-4540,4541-4542,4543-4544,4545-4546,4547-4548,4549-4550,4551-4552,4553-4554,4555-4556,4557-4558,4559-4560,4561-4562,4563-4564,4565-4566,4567-4568,4569-4570,4571-4572,4573-4574,4575-4576,4577-4578,4579-4580,4581-4582,4583-4584,4585-4586,4587-4588,4589-4590,4591-4592,4593-4594,4595-4596,4597-4598,4599-4600,4601-4602,4603-4604,4605-4606,4607-4608,4609-4610,4611-4612,4613-4614,4615-4616,4617-4618,4619-4620,4621-4622,4623-4624,4625-4626,4627-4628,4629-4630,4631-4632,

Db 786 ---GGEKPM-TEAERQSLFPKKVEKWDIP-PEKTVOQVQDKICEWKTYSRPNAKI 840  
Qy 1382 TWYKDGKLSSSS---KYRMEAVGCTRLRVVQOACQADTGEYSCEAGQRLSFLDVAEP 1438  
Db 841 RMYKORKEIFSGGLKYKIVIEKNVCT--LIINNEVDVDTGKYTCANGVPHQAQLTVLEP 898  
Qy 1439 KVFPAKEQVUREVQOAGASTTILSCEVAQAOTEVMYKOKKLSFSSK---VRMEAVG- 1494  
Db 899 PMKYSFLNPLENTQEIYRTKQAVLTCKVNTFRAPLVYRGSKAIOEGDPRFIIEKDAVR 958  
Qy 1495 CTRRLVQOQADAGEYSCEAGSQRLS-FHLHVAEPKAVFAKEQBPASREVOAEGATSAT 1553  
Db 959 CT--LTIKEVEDDOAEWTARILTQDVFSKVQVYVEERHPTFVPMKSKQKNESDL---AT 1013  
Qy 1554 LSCEVAQAOTEVTVYKOKKLS--SSSKVRMEAVGCTRLRVVQOAGADAGEYSKADQR 1612  
Db 1014 LETDVNDKDAEVMWHDGKRIDIDGVKPKVSESSRKRLIINGARIEDHGEYKCTTDDDR 1073  
Qy 1613 LSFHLHVAEPKVPVPAKEQ--PAHREVOAEAGASATLSCEVAQAOTEVTVYKOKKLS-- 1668  
Db 1074 TMAQI-----IVAKKFIVALKDTVEIKDDVTLKQYKDTKIPGIWFRANGKQISSMP 1127  
Qy 1669 SSKVRMEAVGCTRLRVVQOQADAGEYSCEAGQRLSFRHLHVAEL--FOISERPCK-- 1724  
Db 1128 GKFETQSRNGTHTLKIGKIEWNEADVYIEDQAGLRGSCNVTVLEAKRPILANWKPKKIE 1187  
Qy 1725 ---REPLVVKHEHDIILTATLATPSAATVWLKGVIRSKRH--ETASQGDHTLTVH 1779  
Db 1188 AKAGEPCVVKVFFQIKGTRR-GDPKAQI---LKNKGPIDEMRKLEVEIKDDVAEIVFK 1243  
Qy 1780 QAQVLDIAIYSCRVG-----AEGQDF-----PV 1802  
Db 1244 NPQLADTKWALELGNAGTALAPFELPVKDKPKPKPLTKNTVAGGLDVLWGTDPDP 1303  
Qy 1803 QVEVAAKFCLLE---PVCBELGTVTLACELSPACABVVR-----CGNTOPRV 1850  
Db 1304 EGAPYKAYIIEQEGRSNGNAKVGETKGTDFKVKDLXEH-GEYKFRVYKALNECGLSDPLT 1362  
Qy 1851 GKRFORMAEGPVRSUTVLGLRAEADAGEVYCVSRDDHTSAQLTVSVPRVFMKSLSTWA 1910  
Db 1363 GE--SVLAKNP-----YGVFGKPKMDAIDVDKDHCT-----LAMEPP 1398  
Qy 1911 EGGBATFOCVSPSDVAVWFRDQALLOPSEK-----FAISQSGASHSTISDLVEDA 1965  
Db 1399 BEDGGAPI-----GYIIEREKSEKWHQVGTQPCCELTDKKVVEDK 1443  
Qy 1966 GOI-TVEAGSSSAAALVREAPVLFK-KKLEPO-----TVEERSSVTLEVELT-R 2013  
Db 1444 EYLRYVKAVNKAAGPDCDHGKPIKOKAKKASPEFTGGGIKDLRKVGETIKYDVPISGE 1503  
Qy 2014 PMPELRWTRNATALAPGKNVIEHAEAGRHRLVHNVGFDARGFF-----GCETPDD 2064  
Db 1504 PLPECLWVVGKPLKAVGRVMSRSGKHIMKIENAVRADSGKFTITLKNSSGSCD--- 1559  
Qy 2065 KTOAKLTVMROVRLVGLQAVEAREQGTATVE-----VOLSHADVGS 2108  
Db 1560 -STATVTVGVRTPPKPGLDIADVCADG-ATLSMWNPPDDGGDFLTGIVBEAQDMNKGK 1617  
Qy 2109 WTRDGLRFOQGPTCHLAVRGPMTLTLGSLRPEDSGLMVFAEGVHTSARLVWVTELPVSV 2168  
Db 1618 YIEVG-----KVDPTNTLLKVNGLR--NKGNYKFRVAVKANNGE----- 1654  
Qy 2169 SRPLO-DVVYTEKEKVTLECELSRP-----NVDVRW-----LKDQ-----VELRAG 2208  
Db 1655 SEPLSADQYQIKDPWDEPGKPRPEITDFADRIDIAWEPPHKGAGAPIBEYIVVEVRDP 1714  
Qy 2209 KTMALAAQACBSLTIYCEPADQGVYVCDADHAGSSASVKVQGTYYTLIYRRVLAEDAG 2268  
Db 1715 DTK-----ENKE-----VKRPDPTNWSISGLKEGKEVQFRVRAVNKAGPG 1754  
Qy 2269 EIQFVAENAESRAQL-RVKELPVTILVRPLRDKIAMERHGVLECOVSRASA----- 2318

Db 1755 Q-----PSEPSKQJLAKPKFIPAWL-----KHDNLKSITVKAGATVRWEVKIGG 1798  
Qy 2319 ---QVRWPKGQELQPGPKYELVSD-----GLVRKLIISDVHABD 2355  
Db 1799 EPIPEVKWFKGNQQLENG--IQLTIDTRKNEHTILCIPSAMRSDVGEYR-LTVKNSHGAD 1855  
Qy 2356 E-----DVTCTDAGDVKTSQAQFVVEQSITIVRGLQDVTVMBP-----AP 2395  
Db 1856 EKANUTVLVDRFSKNGPLEVSDVF---EDNUNLSWKPPDDDDGGGPIEYVEVEKLDATG 1912  
Qy 2396 AMFEC-----EFSI 2404  
Db 1913 RWVPCAKVKDTHKAHIDGLKKGQTYQFRVAVNKEGASDALSTDKTKAKNPYDEPKTGT 1972  
Qy 2405 PSV-----RPPKWLIGKTVLQ-----AGNVGLE-----Q 2429  
Db 1973 PDVVDWADRVLSLEWEPKSGDGAFTQVIEKKKGKHDWQEOCKVSGDQNAEILGUK 2032  
Qy 2430 EGTVHMLRLRTCTMTGPHFTVGKSRSSARLVVSDIPVILTRPLEKPTGRELQSVVL- 2488  
Db 2033 EGEEYQFRYKAV--NKAGP-----GEASDPSKVVAK-----PRNLKPMIDREAMKTI 2080  
Qy 2489 ---SCDF-----RPAPKAVOMYKDDTPLSPSEKFKMSLEGOMAEHLRLILMPADAGVY 2538  
Db 2081 KVGNDVFEFDPVVRGEPKPK-EMIFNEKPDV-DQKIRIESEDKTRFVLRGATRKHAGLY 2138  
Qy 2539 RCOA-----GSAHSSTVTVAREVTVTGLODAEATEGWAGSFSCELSHEDBEVMSL-- 2592  
Db 2139 TLTATWASSDXHSVEVILGKPSPLGLEVSNVY-----DRADLEWKVPE 2186  
Qy 2593 --NGMPLYNDSPHETSHK-----GRRHTLVLSIORADAGIVRASSLKVSTSRAL 2640  
Db 2187 DDGGAPI---DHYEIEKMDLATGRWVPCGRSET-----TKTTPVNLQPGHEY 2230  
Qy 2641 EYRVKPVVFLKALDLSAE---ESGTLALQCEVSDPEAHVYVRKDGVOVG- 2687  
Db 2231 KFRVAVNKEGSDPLTTNTAILAKNPYEVPGKVDPKPEL-VWMDKDHVDLANNAPDDGA 2289  
Qy 2688 ---PSD-----KYDFLHTAGTGLVHVHDVSPED 2712  
Db 2290 PTEAFVIEKKNGRWEALVVGQDKTATVNLKEGEBYQFRISARNKAGTGDPSDPSD 2349  
Qy 2713 AGLYTCHVGSSETRAVRVVHDHLVHGITKXLTMEVLEGESCFEVLSHESADPAMTV 2772  
Db 2350 ---RVVAKPRNLAPRIH-----REDLSDTTVKVGATLKFIVHIDGEPAPD-VTWBF 2396  
Qy 2773 GGTIVGSS-----SRF-----OATQRGKYLIVVREAAAPSADAGEVVFVSRGLTSK- 2817  
Db 2397 NGKIGESKAOIENEPYISRFALPKALRKQSKYITATNINGTDSVTINIKVSKPTKP 2456  
Qy 2818 ---ASLIVRERPAIIPLEDQ-----WV----- 2838  
Db 2457 KGPIETDVFEDRATLDWKPDPEDDGGEPIEFYIEIKONTKGIWVPCGRSGDTHFTVDSL 2516  
Qy 2839 ---APGEDVELRCEL-----SRAGTP-----VHV--- 2859  
Db 2517 NKGDHYKFRVAVNSEGSDPLETETDILAKNPPRDPGPEPTDMDSDHVDLKNDDP 2576  
Qy 2860 LKDRKA-----IRKSQYD-----VVCETWAMLVIRGASIKDAGEYTCV----- 2900  
Db 2577 LSDGGAPIBEYQIEKTKYGRWEPAITVPGQTTATV--PDLTNPBEYEFVRVAVNKGGP 2634  
Qy 2901 ---EASKSTASHVEEKANCANTEELTNLOVEKGTAVFTCKTE-HPAATVTVWR-KGLLE 2954  
Db 2635 SDPSDASKAVIAKPRNLKPHIDRDALKNLITIKAGQISIFDVPGSGPATVTVHWNDR 2694  
Qy 2955 LRASGHQPSQEGTLURLTISALEKADSDTYTCIDIGQAQSRQAQLLVQGRVHIE----- 3009  
Db 2695 IRNGRKLNDNPYQSKLVVKMERGDSGFTTIKAVNANGEDBATV---KINVIDKPTSP 2751  
Qy 3010 ---DLEDVD-----VQEGSSAT----- 3023  
Db 2752 NGPLDVSVDVHGDVTLNWRAPDDDDGIP-ENYVIEKYDTASGRWVPAKVAGKTTAVVD 2811

QY 3024 -----PRCRISPANYE-----PVHWF-----LDKTPLE-----HANELNEID----- 3054  
 Db 2812 GLIFGHEYKFRVAANVAGESDPLETFGTTLAKDPFKPKTKTNAPEITDWDKOHVDLEWK 2871  
 QY 3055 --AOPGG-----YHVLTLRLQA-----LKDSGTIYF-----EAG 3081  
 Db 2872 PPANDGGAPIEEYVVMKDBSPFNDVAHVPAQGTNATVGNLKEGSKYEFRIKAKNAG 2931  
 QY 3082 ----DQASAAALRTEKPSVFSR-ELTDATITEGEDLTVCETSTCDIP-MCWTKOGKTL 3135  
 Db 2932 LGDPSDASAVAKARNPPVVDIENSIQBIKVKAGQDFSLNIPVSGEPTTITTFEGTPV 2991  
 QY 3136 RGSARCOLSHBGHRAQLIITGATIQDSGRY--KCEAGGACSSSIVRVHA-----RPV 3185  
 Db 2992 ESDDRMKLNNEBDGKTKFHVXREALSDGTIYIIKAENENGDTAEVKVTVLDHPSPPRGPL 3051  
 QY 3186 RFQALKDLEVL-----EGGAATLRCVLSSV-AAPVKV-----CYGN 3221  
 Db 3052 DVTNIVKDGCDLAWKEPEDDGGABEISHYVIEKQDAATGRWTACGESKDTNFHVDDLTOGH 3111  
 QY 3222 N-----VLRPGDKYSLRQEGAML-----ELV-----VNLRLPQDSG 3252  
 Db 3112 EYKFRVAVNRHSGSDPLEAREALIAKDPFDRADKPGTPEIVDWDKDHADLKWTTPADGG 3171  
 QY 3253 RYSCSFGDQTSATLVTALPAQFIGIKLRNKEATEGATATLRCELSKTAPVWRKGETL 3312  
 Db 3172 GAPIE-GYLVEMRTPSGDWVPAVTG-----AGE-LTATV-----DGL 3207  
 QY 3313 RDGRYCLRQDGCANCELQIIGLAMVDAEYSCVCGEERTSASLTIRMPAHFIRLEH-- 3370  
 Db 3208 KPGQTY-----QFRVAKLAKAGE-----STPSDPSRTWAKPHLAPKINRDM 3250  
 QY 3371 ---QESIEGATAT--LRCELSKAAPVEW-RKGRESLRDGRHSLRQDGAIVCELOICGLAV 3424  
 Db 3251 FVAQVRKAGQTLNFDVAVGEPAKPIEWFLNGSPLSGCGNTHDNNNTKLTTRKSTAR 3310  
 QY 3425 ADAGEYSCVC---GEERTSATLTKALPAKFTEGLRNEEAVEGATAMLCESLKVAPVE 3480  
 Db 3311 ADSKYLKIVATNBSGKDEHEVDVNIIDIPGAPGRLKHDKITKESVVVLKWE-----PLD 3365  
 QY 3481 WRKGP-----ENLRDGRYILROEGTRCELOICGLAMADAGEY--LCVCGOERTSATL 3531  
 Db 3366 DGGSPITNYVVEKQDGRWVPCGETSDTSIKYNKL--SEGHEYKFRVAVNRQGTSAPL 3423  
 QY 3532 TI-RALPARFIEDVQOAREGATAVLOCELNSAAPVWRKGETL-----RDG-----D 3580  
 Db 3424 TSDHAIIVA-----KNPFDPEFAP-----TDVTPVDMDKHVDLEWKPPANDGGAPID 3470  
 QY 3581 RYSL-----RQDG--TKC-----ELOIRGLAMADTGEYS-----CV 3609  
 Db 3471 AYIVEKKDKFGDWVECARVDGKITKATADNLITGETTYQFRVAVNRKAGPGKPSDPTGNVV 3530  
 QY 3610 CGOERTSAMLTVRA---LPKIFTEGLRNEEATEGATAVLRCELSKMAVWRKGBE-TLR 3665  
 Db 3531 AKPRMAPKLNLAGLLDLRIKAGTPIKLDIAFEG-----EPAPVAKWKANDATID 3580  
 QY 3666 DGRHSLRQDGCARCELOIRGLVAEDAGEYLCMC-----GKERTSAMLTVRAMSKFIEGLR 3721  
 Db 3581 TGARADVTNPTSSAIHIFSAVRGDTGVYKIIIVENEHGKDTAOCNVTLVDVPGTPEGPLK 3640  
 QY 3722 NEEA-TEGDTATLWCELSKAAPVWRKGBETLR-----DGRHSLRQDGRSCELOIRGL 3774  
 Db 3641 IDEIHKEG-----CTLNWKPPPTD-NGGTDVLHYIVEKMDTSRGTQWQEVGTFDCTAKVN 3693  
 QY 3775 AVVDAGEYSCVCGOERTSA-TLTVRALPARFIE----- 3806  
 Db 3694 KLVPGEYAF-----RVKANVQESKPLEAEPIIAKNOFQVDPDVPDKPEVTDMDKORI 3748  
 QY 3807 DVK-NOBAREGATAVLO--CELSKAAPVWRKGETLGRGDYSL--RODGTRELQIIGH 3861  
 Db 3749 DIKWNPTANNGGAPVTGYIVEKKEKGSIAIWEAGKT--PQTTESADNLKPGVEYEFVIA 3806

QY 3862 LSVADTGEYSCVCGOERTSATLTVRAPOFVFPREPLOSLQAEEGSTATLQCE-LSEPTATV 3920  
 Db 3807 VNAQPSDPSDPTDQIITKA-----RYLAPKILITASKIKIKAGFTHNLEVDIFGAPDDTA 3862  
 QY 3921 VMSKGLQLOANGREPRPL-----QCGTAELVLQLOREDTGEYTCCTCGSQATSATLTVTA 3976  
 Db 3863 TWTVG-----DSGAALAPELLVDKASTSTTSIFPSAKRADSGNY----- 3901  
 QY 3977 APVRFELRLOHE-----VDEGGTAHLCCSLSRAGASVEMRKSGLQLPFPCAKYQMV--- 4027  
 Db 3902 -KLKYNELGEDEALFEVIVQDRPSA-----PEGLEVSDVTKDSCLVNW 3945  
 QY 4028 ---QDGAALLLVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFKTRLOLSQETGDI 4084  
 Db 3946 KPPKDDGGAGELSNVYVVEKRDITNTW-----VPVSFAVT----- 3979  
 QY 4085 ARLCCQLSDAESGAVVQWLKEGVELHAGPKYEMRSQGTATRELLIHOLEAKDTGEYACVTG 4144  
 Db 3980 -----GTSITVPKLTEGHE-----YEFR---VMAENTFGRSDSLNTDFVLAKD 4020  
 QY 4145 GOKTAASLRVTEPEVTIVRGELVDAEVTADDEVEFS-----CEVSAGATGV 4190  
 Db 4021 PFGTPE--KPGPEI-----VD---TDNDHIDIKWOPPRDNGGSPVDHYDIERKDAKTG 4069  
 QY 4191 QWC-LOGLPLOSNEVTEVAVRDGRIHTRLKGVTPEDAGTVSFHLGNHASSAQLTVRAP 4249  
 Db 4070 RMIKNTSPVQGTAFSDTRVQKG--HTVEYRVVAVNRKAGPQPSDSSAAAATAKPMHEAPK 4127  
 QY 4250 VTILLEPQDOVOLSEGODASFOCLSRASGQEARWAL--GGVPELOANEMNDITVEOGLHL 4308  
 Db 4128 FDLDDGHEFRVKAPELIVITPTASPOPDISHTKEGKPLAGVETDTSQTK-----L 4181  
 QY 4309 TLHKVLTLEDAGTVSFHGT--CSSEAOQKVTAKNTVVRGLNVLENVALEGGEGALFECQLSQP 4366  
 Db 4182 VIPSTRRSDSGFVKIKAVNPYGEAEANIKITV-----IDKGAPE-----NITYP 4226  
 QY 4367 EVAHTWLLDDEBPVRTSENAEVVFENGRLHLLKLNLRPODSCTVTFIAGDMVTSFLT 4426  
 Db 4227 AVSRHTCLNWDAPKDDGGAIEAGTKIEYQEV-----GQIKWKPGLISGTAYT 4276  
 QY 4427 VRGWRLEILEPKNAARVAGAAQAFCTTSLBAVPVGEASWYINGAAV---QPDSDSDTWT 4483  
 Db 4277 VRG-----LEHGQYFRIRAEANAVGLSD--YCQGVPPVVIKDPDP----- 4315  
 QY 4484 ADGSHQALLRSQAHPHAGEVTFACRDAVASARLTVLGLDPPDEAEVVAHSHSTVILSW 4543  
 Db 4316 -----PCAPSTPEITGYDTNQVSLAW 4336  
 QY 4544 AAPMSDGGGGLCGYRVE--VKEGATGQWR-LCHELVGPECVVDGLAPGTFPRVAAVG 4600  
 Db 4337 NPFRDDGGSPILGYVVERFEKRG--GDWAPVEMPVWKGTECIVPLGHNENITYQFRVAVN 4395  
 QY 4601 PVGAGEPVHLPTQVRLAEPKPPQPSAPESROVAAAGEDVSLSELEVVABEAGEVWHKGM 4660  
 Db 4396 AAGHEPSNGSEPVTC---RPYVEKPGADPAVRG-----KITQSAELTMNR-- 4440  
 QY 4661 ERIOPGG-----RFEVVUSQGRQOMLVIKGFTABDQ 4691  
 Db 4441 -PLRDGGAPIDGYIVEKKLGDNDWTRCNDKPVDRDTAFEVKNLGEK----- 4486  
 QY 4692 EYHCGLAAGSTCPAAATFQVALSPASVDEAPQPSLPPEAAQOEGDLHLLMALAKRMSR 4751  
 Db 4487 EYEFV-----IAVSAGEGEPKPS-----DLVIEQPGR----- 4518  
 QY 4752 EPTLD--SISELPEEDGSR--QRLPQAEAEVAP--DISEGVSTADELART-----CDAD 4799  
 Db 4519 -PIFDINLKDITVRAGETQIRIPYAGGNPKPIIDLFGNSPIFENERTVVDNPGEIV 4577  
 QY 4800 LSHSSDDESAGTSPSLVTLKKAGR-----PG--TSPL-ASKVGAPAAP-SV 4843  
 Db 4578 ITTGS-KRSDAG-PYKISATNKYKDKTCLNVFVLDAPGKPTGPIRATDIQADAMTLSW 4635  
 QY 4844 KPQQOQ-----EPLAAVRPPLDGLSTKDLGDPMSMOKAAVKIQAAFKGVKVRKEMKQEG 4897

Db 4636 RPPKNGGDAITNVVBEKTRPGGDWVT--VGHVPVGTTLVRNLDANTPYBFRVRAENQYQ 4693  
 QY 4898 PMFSHTFGDTBAQVGDALRL-----ECWVASKADVFAARWLK---DG 4935  
 Db 4694 -----VGEPLETDDAIVAKNPFDTGAPGQPEAVETSEBAITLQWTRPTSDG 4740  
 QY 4936 -----VE-----LTDGRHHHIDQGGTCSLLIAGLDRADAGCYTC 4971  
 Db 4741 GAPIQGVIEKREVGSTEWTKAAFGNILDTKHRVTGLTPKTYEFRVAAYNAAGQGEYS- 4799  
 QY 4972 QVSNKFGVTHSACVWVSGSEASESSGGELDDAFRAARLRLFR-TKSPAENVSDDE 5030  
 Db 4900 -----VNSVZITADNAPTRKINNGMLTRDILAYAGERAKILVPPFAASA- 4846  
 QY 5031 LFLSADGAPBEPBPAQWQVREDEHFCIRFEALTEARQAVTFQMFATLIGIVEIKL 5090  
 Db 4847 VTFSGKNGKISPTDRVKVEYSPLATLTIKSELTDGLYFVELENSQSGDSASIRLKV 4906  
 QY 5091 VEQ--GRRVEMCISKETAP-----VVPPEPLPSLLTSDAPV--FLTELQNEVQDG 5140  
 Db 4907 VDKPASQHIRV--EDIAPDCCTLYWMPSS--SGGSPITNYIVEKLDLRLHSDG 4956  
 QY 5141 -----YPVSFDC-----VVTGQP- 5153  
 Db 4957 KWEKVSSVVRNLNVTGGLIKDNRYRFRVRAETQYGVSEPCELADVVAVKYQFEVNPQPE 5016  
 QY 5154 MPSV-----RWFKDGKLEBDDHYMINEDQ 5178  
 Db 5017 APTVRDKDSTWAELEWPPRGGSKIIIGYQVQVADTSSGWINAKMDLSQCHARTGLR 5076  
 QY 5179 QGGHOLIITAVPADMGVYRCLAENMGVS--STKAELR-----VDLTS-- 5220  
 Db 5077 QNGE-----PEFRIIAKNAAGFKSPSPSERCQKSRFGPPGPIHVGAKSIG 5124  
 QY 5221 -----TDYTDADATSSSVFSAGGLSSREQGTESTTDE----- 5256  
 Db 5125 RNHCTITWAPLEDGSKITQYNEIREYGSTLMTWASDY-NVREPEFTVDKLRREFNDYE 5183  
 QY 5257 -----GQLPQVVEELRLQVAPGTGLAKFQLKVGKGYPA 5289  
 Db 5184 PRVAINNAGKIPSLSPGKIKIQUESGSRPQIVVKPED-TAOPYNRAVFTCEAVGRPE 5242  
 QY 5290 PRLYFKDQGPLTASAHIRMTGKXILHLEIISVTRDSQGYAAYISNAMAAYSAARLL 5349  
 Db 5243 PTARWLRNGRELPESSRYRFEASDGVYKFTIKEWMDIDAGEYTVESNPGYSDTANLV 5302  
 QY 5350 VRGPEDEPEKASDVHQLVPPRMLERFTPKVKKKGSSITFSVKVEGRPVPTVHVL---R 5406  
 Db 5303 VQAPPVIEK-----DVPNTILP-----SGDLVRLKIYFSG-TAPFRHSLVLNR 5344  
 QY 5407 EEABRGVLWIGPDTFGYTVASSAQOHSILVLDVGRHQGTYTCTIASNAAGQALCSASLHV 5466  
 Db 5345 EE-----IDMDHPTIRIVEFDDHILITIPALSVREAGRYEYTVSNDSGEATTGFWLNV 5397  
 QY 5467 SGLPKVBEQKVEALISTFLQGTQTAISAOGLTASADLGGORKEEPLAAKEALGHS 5526  
 Db 5398 TGLPEAPQ-----GPH 5409  
 QY 5527 LAEYTBEBFLQKTSQITEMVSAKITQAKLVPGGDSDESKTPSAPRGRSRPSSIIQ 5586  
 Db 5410 ISNIGPSTATLSWPPVTD-----GGSKITSYVV 5438  
 QY 5587 ESSSESEDGDARGEIFDIYVVVADYLPGLAQDAITLREQQYVEVLDAAHPLRLVTKP 5646  
 Db 5439 EKRLSKD-----EWVTVTSNVKDMNYIVT--- 5463  
 QY 5647 TKSSPSRQGVSPAYLDERLKLSPWGNAAEAPFPGEAVSEDEYKARLSVIOQLLSSEQ 5706  
 Db 5464 -----GLFE-----NHEYEFVSA----- 5477  
 QY 5707 AFVBELOFLQSHLQLHLERCCHVPIAVAGQKAVIFRNVDRIGRPHSSFLOEQCUDDDD 5766

Db 5478 -----QNENGIG----- 5484  
 QY 5767 VAMCFIKNOAAFEQYLEFLVGEVQAESVWVSTAIQEFYKXKAAEEALLAGDPSPQPPPLQ 5826  
 Db 5485 -----APLVS-----EHPITIALRPPDPPTSPL- 5506  
 QY 5827 HYLEQPVVERVORYQALLKELTENKARNQNCALLBOAYAVVVSALPQRAENKLHVSIMENY 5886  
 Db 5507 -----NLEUVQGGDYVTLWSOR----- 5524  
 QY 5887 PGTLEALGEPRIQGHFIVMEGAPGARMKPGHNRHVFLEFRNHLVICKPRRDSRTDTSYV 5946  
 Db 5525 -----PLSDG-----GGRLL-----RGYIV--EKQEEHDEWFR 5551  
 QY 5947 FRMMKLSIDLNQDVEGDDRAFEVQEREDSVRXLYLQARTAIKXSWWKEICIQORL 6006  
 Db 5552 NQNPSPNNYVNPILDG-----RKY--RYRVAYNDAGLSLAELODTL 5594  
 QY 6007 --ALPWRPDPPEBELADCTABLGTVTKLACRVTGTPKPVTSWYKDGKAVQVDPHHILIE 6064  
 Db 5595 FOASGSGEGPKIVSPSLDNEEVGRCVTFECEISGSPREYRNFKGCKELVDTSKTYLIN 5654  
 QY 6065 DPGSCALILDSTGVDSQYMCFAASAAGNCSTLGLKLVQ-----VPPRFVNVKVRASP 6118  
 Db 5655 KGDKQ-VLIINDLTSDADDEVTCRATNSGSTRANLRIKTKPRVFIPPKYHGYEAQ- 5712  
 QY 6119 FVEGEDAQFTCTIEGAPYQIRWYKDGALLTTGNKFQTLSEBSGLLVIVRAASKEDLG 6178  
 Db 5713 --KGETIELKIPYKAYPOGEARWTKDGEKIENNSKFSITDDK--FATLRSNAREDYG 5768  
 QY 6179 LYECELNRLGSRASAEIRIQSPMLAQEQCHREQLVAVEDITTLERADQEVTVLKL 6238  
 Db 5769 EYRVVSVNSVSGSDGTANVTV----- 5789  
 QY 6239 LGPKAPGSGTDLTGPGCPGCPALQBTGSPVVTGTSEAPVPRVP--QPLLHEGP- 6295  
 Db 5790 -----ADVEP-PRPFITENILDEAVI 5810  
 QY 6296 -EQEPEAI---ARACWTVPIR-MEGAAMPAGAGTCELLMDVHSHVVRETQRTTYQYQID 6350  
 Db 5811 LSNKFPALDGGSLVNTYITEKEANGGSW-----SPCAKSRITYTYTIE 5853  
 QY 6351 -THTARPSMQVTIEDVQAQT---GGTAQFEAIIISGDPQSPSVTWYKDSVOLVDSTLSQQ 6406  
 Db 5854 GLRAGKQYEFRIIAENKHGQSKPCPFTA--PVLIFGDERKRRRGYD---VDEQKIVR 5906  
 QY 6407 QEGTYSILVRHVASKDAGVYTCLAQNTGGQVLCRAELLVLGGDNEPDSEKQSHRKLHS 6466  
 Db 5907 GKGTVSS-----NYDNYVFDIWKQY-----PQPEIKHDX---VLD 5940  
 QY 6467 FYEVKEEIGRGVFGFKRVQHKGNKILCAAKFIPLRSTRQAQYR-ERDILAAALSHPLVT 6525  
 Db 5941 HYDIHEELGTAGFVVHVRTERATGNFPAKFVMTPHESDKETVRKE-QTMSVLRAHPTLV 6000  
 QY 6526 GILDDOFEIRKTLILLILELCSSEELDRLY-RKGVVTEAEVKYIOQLVEGLHYLHSHGVL 6584  
 Db 6001 NLHDAFEDDNNVMVIEYFWSGGLPEKFEKVADEHNKMSDEAVEYMRQVCKGLCHMENNIV 6060  
 QY 6585 HLDIKPSNLMVHPAREDIKICDFGFAQNTITPAELQFSQYSPSEFVSPEIIOQNPFVSEAS 6644  
 Db 6061 HLDLKPENIMFTKRSNELKLLIDFGLTAHLDPKQSVKVTGTGTAEFAAPEVAEKGKPVGYT 6120  
 QY 6645 DTWANGVSYLSLTCSPPAGESDRATLLNVLEGRVSWSSPWAH--LSEDAKDIKATL 6702  
 Db 6121 DWSVGLSYILLGSLPFGGNDDETLRNVKS--CDWNMDDSAFSGISEDGDKDFIRKL 6178  
 QY 6703 QRAPQARPSAOCCLSHPWFLK-SMPAEBAHFINTKOLK 6739  
 Db 6179 LADPNTRMTIHOALEHPMLTPGNAPGRDSQIPSSRYTK 6216

RESULT 9  
 A88852



Db 1307 MQEGRSNWAKVGETKGTDFKVXDLKEH-GEYKFRVKALNECGLSDPLTGE--SVLAKNP 1363  
Qy 1862 VRSLTVLGLRAEDAGEVCSRRDDHTSAQLTVSVPRVVKFMSGLSTVVAEGEATFCQV 1921  
Db 1364 -----YGVPKPKNDMAIDVDHDCT-----LAWPEPDEGGAPI--- 1398  
Qy 1922 VSPSDVAVWFRGALLOPEK-----FAISQSGASHSLTISDLVLEDAGQI-TVAEAGA 1975  
Db 1399 -----TGIIIREREKSEKWHQVGTQKFPDCCCLTDKKVVEDKXYLYRVXAVNK 1446  
Qy 1976 SSSAALVRBAPVLFK-KKLEPQ-----TVEERSSTVLELT-RPWPELRWTRNA 2024  
Db 1447 AGPDPCDHGKPKIMAKKASPEFTGGIKDLRLKVGETIKYDVPISGEPLEPCLMWVNG 1506  
Qy 2025 TALAPGKVIHAEGARHLVNLVNGPADRGFF-----CCEPDDKTKQAKLVEMR 2075  
Db 1507 KPLKAVGRVKSSEGRGHKHIENAVRADSGKTTIILKSSGGSCD-----STATVTVGR 1561  
Qy 2076 QVRLVRGLQAVEAREQGTATME-----VOLSHADVDGSWTRDGLRFQOG 2119  
Db 1562 PTPPKGFLDIADVADG-ATLSWNPDDDDGGDPLTGYVIEAQDMNKGKIEVG----- 1614  
Qy 2120 PCHLAVRGPWHTLTSLGAPEDSGLMVFKABGVHTSARLVVTELPVFSRPLQ-DVVIT 2178  
Db 1615 -----KYDPNTTLKVNGLR--NKGNYKFRVKA VNEGE-----SEPLSADQYTO 1657  
Qy 2179 EKEKVTLECBLSRP-----NVDVRW--LKDG-----VELRAGKTMIAAQQGAC 2219  
Db 1658 IKOPWDPGKPRPEITDFADRIDIAWPPHKGDPGAPIEYIVVEDRDPDK----- 1709  
Qy 2220 RSLTIYCEPADQGVYCDADHQAQSSASVKVQGRYTLIYRVLADAGBIQVAVNAES 2279  
Db 1710 -----EWKE-----VVRPDTNASTISGLKEKEYQFRVAVNKAQPGQ-----PSEPS 1752  
Qy 2280 RAQL-RVKELPVLTVRPLDKIAHEKRGVLEQCVSRASA-----QVRWFKG 2325  
Db 1753 EKQAKPKFIPAWL-----KDNLKSITVKAATVRWEVKIGGPIEYVFKWFG 1801  
Qy 2326 SQELQPGKVELVSD-----GLYRKLIIISDVHAED-----DT 2358  
Db 1802 NOQLENG-IQLTIDTRKNEHTILCIPSAMRSDVGEYR-LTVKNSHGADBEKANLATVLD 1858  
Qy 2359 YTCADGVKTSQAQFVFEZQSIITVIRGLQDVTVNEP-----APAWFEC----- 2400  
Db 1859 PSKPNGLVSDVF---EDNLNLSWKPPDDGGEPYBYEVEKLDATGRWVPCAKVOT 1915  
Qy 2401 -----KAHTDGLKKGQYQFRVAVNKEGASDALSTDKDTAKNPDYBPGKTGTPDVVDWDADR 1975  
Db 2408 -----RPPKWLIGKTVLQ-----AGNVGLE-----QEGTVHRLMLR 2440  
Qy 1976 SLEWPPKSDGGAPITQYVIEKKHGRDWQECQKVSQDQTNABILGLKEGEYQFRVKA 2035  
Qy 2441 TCSTMTGPVHFTVSKRSSARLVSDIPVVLTRPLEPKTRELQSVL-----SCDF--- 2492  
Db 2036 V--NKAGP-----GEADSPKSVAK-----PRNLKFWIDREAMKITIKVGNVDFDVP 2083  
Qy 2493 ---RPAPKAVQYKDDTPLSPSEKFKMSLEGOMAEILRLMPADAGVYFCQA-----GSA 2545  
Db 2084 VRGEPKPK-EWIFNEXPVD-DQKIRIESDYKTRFVLRGATRKHAGLTLTATNASGSD 2141  
Qy 2546 HSTTEVVEAREVTVTQLOABATEEGWASFSCELSHEDEVEWSL-----NGMPLYNDS 2601  
Db 2142 KHSVEVILGKPSPLGPLEVSNVYE-----DRADLEWKVPDDGGAPI---D 2186  
Qy 2602 FHEISHK-----GRRHTLVLSQIRADAGIVRASSILKYSTSARLEVRVKPVVFLK 2651  
Db 2187 HYEIEKMDLATGRWVPCGRSET-----TKTTPNLPQPGHKEYFRVAVNKEG 2233  
Qy 2652 ALDILSAB-----ERGTLALQCEVSDPEAHVVRKDGVLQJ----- 2687  
Db 2234 ESDPFTTNTAILAKNPYEPGVKDKPEL-VWDKDXHVDLAWNAPDDGGAPIEAFVIEBKD 2292

Qy 2688 -----PSD-----KYDFLHTAGTRGLVVDVSPEDAGLYTCHVGE 2723  
Db 2293 KNGRWEELVVPDQOKTATVNLKEGEYQFRIGARNKAGTGDPSPDSD-----RVVAK 2346  
Qy 2724 ETRARVRVHDLHGVTIKELKTMVELEGESCFECVLSHESASDPAMTVVGKTVGSS--- 2780  
Db 2347 PNLAPRIH-----REDLSDTTVKVGATLKFVHIIDGEPAPD-VTWSFNKGIGESKAQ 2399  
Qy 2781 -----SRF-----QATQGRKYILVREAAAPSAGEVVFVRGLTSK-----ASLIVR 2823  
Db 2400 IENEPYISRFALPKALRKQSGKYITATNINGTDSVTINIKVSKTKPKGPPIEVDVFE 2459  
Qy 2824 ERPAAIKPLEBQ-----WV----- 2838  
Db 2460 DEATLWKPPEDDGGEPYEFYIEKMTKDGWVPCGRSGDTHFTVDSLNKGDHYKFRVK 2519  
Qy 2839 -----APGSDVELRCEL-----SRAGTP-----VHW-----LKORCA----- 2865  
Db 2520 AVNSGSPDPLETETDILAKNPFDRPDRPPEPTDMDSDHVDLKWDPPLSDGGAPIBEY 2579  
Qy 2866 -IRXSQKYD-----VVCETMAMLVIRGASLKAGEYTCV-----EASKSTA 2907  
Db 2580 QIEKTKYGRNEPAITVGGQTTATV--PDLTPEEYEFVRVAVNKGGPSDFPSDASKAVI 2637  
Qy 2908 SLHVEEKANCFTTELTNIQVBEKPAVFTCKTE-HPAATVTVR-KGLELRASGHKQPSQ 2965  
Db 2638 AKPRNLKPHIDRALKNTIKAGOSISFDVPVSGEPAPTVMHPDNRREIRNGRVLKDN 2697  
Qy 2966 EGLTURLIISALEKADSDTYCDIGQAQSRALVLQGRVHLIE-----DLEDVD- 3015  
Db 2698 PEYQSLVVKOMERDGSFTTIKANVANGEDATV---KINVIDKPTSPNGPLDVSVDHG 2754  
Qy 3016 -----VOEGSSAT-----FRCR 3027  
Db 2755 DHVTLNWRAPDDGGIPIENYVIEKYDTASGRVPAKAVAGKTTAVVDGLIPGHEYKFR 2814  
Qy 3028 ISPANYE--PVHWF--LOKTP--HANELNEID-----AQPG----- 3059  
Db 2815 VAAVNAEGESDPLETFTGTLAKDPXPKTNAPETDMDKDXHVDLEWKPPANDGGAPIE 2874  
Qy 3060 -----YHVLTLRQLA---LKSGTIFY-----EAG-----DORASAA 3088  
Db 2875 EYVEMKDEFSPFNDVAHVPAQTNATVGNLKGSKYEFRIACNKAAGLGDPPSASAV 2934  
Qy 3089 LVTEKPSVFR-ELTDATITEGEDTLVCTETSCDIP-MCWTKDGKTLRSARCOLSHE 3146  
Db 2935 AKARNVPPVIDRNSIQEIKVKAGQDFSLNIPVSGEPTTITWTFFEGTVESSDDRMKLNE 2994  
Qy 3147 GHRAQLLITATLODSGY--KCEAGGACSSSIVRHA-----RPVRFOEALKDLEV 3196  
Db 2995 DGKTFHVRALRSDTGYIIEKENENGTDIAEVKTVLDHPSSPRGFLDVTNIVKDGCD 3054  
Qy 3197 L-----EGGAATLRCVLSSV-AAPVKW-----CYGNN-----VLR 3225  
Db 3055 LAWKEPDDGGAIEIHYVIEKQDAATGRWTAGESKDTNFHVDDLTQGHYKFRKAVNR 3114  
Qy 3226 PGKYSILRQEGML-----ELV-----VRNLRPQDSGRYSCSGDQTT 3263  
Db 3115 HGSDFPLEAREAITAXDPFDRADKFGTPEIVDMDKXHADLKWTPPADGGAPIE-GYLVE 3173  
Qy 3264 SATLTVTALPAQFIGLKNKEATEGATATLRCELSTAPVEWRKGSETLRDGDYCLRQD 3323  
Db 3174 XRTPSDWMVPAVTG-----AGE-LTATV-----DGLKPGQTY----- 3205  
Qy 3324 GAMCLOIRGLAVMDAAEYSCVCGEERTSASLTIRMPAHFIRGRH-----QESIEGAT 3378  
Db 3206 -----QFRVKALNAGE-----STPSPRSTWVAKPRHLAPKIRNDRMFVAVRKAQOT 3253  
Qy 3379 AT--LRCELSKAAPVEM-RKGRESLRDGRHSRLQDGAIVCELOICGLAVADAGEYSCVC- 3434  
Db 3254 LNFQVNEGEPAPKLEWFLNGSLSSGNNTHIDNNTDNTKTTTSTARADSGKXIYAT 3313





Db 5128 LEDGSGKITGYNVEIREYVSTLWTVASDY-NVREPEFTVDKLRFNDEYFRVAINAAGK 5186  
 QY 5257 -----GQLPQWBEELRLQVAPGTRLAKFQKLVKGYAPARLYWFKDQP 5300  
 Db 5187 GIPSLPSGPIKIQESGSRPQIVVKPED-TAOPYNRRAVFTCEAVGRPEPTARLWRNGRE 5245  
 QY 5301 LTASHIRMTGKKLHTLEIISVTRDSGOYAAIISNAGYASABLLVRGPEPEKP 5360  
 Db 5246 LPESRVRFEASDGYKFTIKEDWDIDAGEYTVESNPNYSGDTATANLVQVAPVIBK-- 5303  
 QY 5361 ASDVHEQLVPPMLERFPKKVKGSSITFSVKVGRPVPTVHWL---REEAERGVLWIG 5417  
 Db 5304 --DVPNTILP-----SGDLVRLXIYFSG-TAPFRHSLVLNREE-----ID 5340  
 QY 5418 PDTGQYTVASSAQHSLVLLDVGRHQOQTYTCIASNAGQALCSLHVSGLPKVBOEK 5477  
 Db 5341 MDHPTIRIVFDDHILITIPALSREAGRYEYTVSNDSGEATTFWNLVTGLPEAPQ--- 5397  
 QY 5478 VKEALISFLQGTQAIQAQGLTASPADLQGRKEBPLAAKALHSLAEVCTEEFLQ 5637  
 Db 5398 -----GPHLSNIGPSTATL 5412  
 QY 5538 KLTQITMVSAKITQAKLVQPGDSDESKTPASPHGRSRPSSISQSSSESODGA 5597  
 Db 5413 SWRPVTD-----GGSKITSYVVKRDLSDK--- 5438  
 QY 5598 RGEIFDIYVVTADYILPLGABQDAITLREGQVVEVLDAAHPLRLVLRVTRKPTKSSPSRQGW 5657  
 Db 5439 -----EWVTVTSNVKMNIVT----- 5455  
 QY 5658 SPAYLDRRLKLSPEWGAABEPEFGEAVSEDEYKARLSSVTOELLSEQAFVEBLQFQS 5717  
 Db 5456 -----GLFE-----NHEYEFVSA----- 5469  
 QY 5718 HHLQHLERCPHVTAQCKAVIFRNVDRDIGRPHSSFLQELQCDDDDVAMCFIKNQAA 5777  
 Db 5470 -----QENGIG----- 5476  
 QY 5778 FEQYLEFLVGRVQABSVVSTAIQBFYKKAEEALLAGDPQPPPLQHVLEQPVERVQ 5837  
 Db 5477 -----EPIIARLPDPTSP----- 5498  
 QY 5838 RYQALLKELIRKARNRQNCALLEQAYAVVSGALPQAKENKHLVSLMENYPTLEALGEPI 5897  
 Db 5499 -----NLEIVQGGDYVTLWMQ-----PL 5518  
 QY 5898 RQGHFIVWEGAPGEMPMKGNHRHVFLEPHLVICKPRDSRTTTSVYFRNMKLSID 5957  
 Db 5519 SDG-----GGRL-----RGVIV-----EKQEEHDEWFRCNQPPFPNNYN 5554  
 QY 5958 LNDQVEGDDRAFEVWQEREDSVRYLLQARTAIKSSWVKICIGIQRL--ALPVMRPPD 6015  
 Db 5555 VPNLIDG-----RKY--RYRVFAVNDAGLSDLAELDQTLFQASGSGGPK 5597  
 QY 6016 FEEELADCTAELGETVVKACRVGTGPKVPSWYKDGKAVQVDPHHLIEDPDGSCALILD 6075  
 Db 5598 IVSPLSLDNEEYGRVCVTFECSIGSPRPEYRWFKCKELVDTSKYTLINKDKQ-VLIIN 5656  
 QY 6076 SLTGVDGQYMCFAASAAGNCTGLKILVQ-----VPPRFVNVKVRASPFVEGEDAOTFC 6129  
 Db 5657 DLTSDDADEYTCATNSGSTRANLRIKTKPRVFIIPKYHGGYEAQ--KGETIELKI 5713  
 QY 6130 TTEGAPYPOIRWYKDGALLTGNKFTLSEPRSGLLVLVIRAAKEDLGLCEBLVNRIG 6189  
 Db 5714 PYKAYPQGEARWTKDGEKIENNSKFSITDDK--FATLRISNAREDYGEYRVVENSVG 5771  
 QY 6190 SARASAEELRIQSPMLQAQEQCHREQLVAIVEDTLERADQEVTSVLKRLGPKAPGPGTG 6249  
 Db 5772 SDSGTUNVTV----- 5781  
 QY 6250 DLTFGPCPCRGAPALQGTGSPVVTGTSEAPVPRVP--QPLHGEPP--QOEPAI--- 6302  
 Db 5782 -----ADVPE-PPRFPIIENILDEAVILSWKXPALDGG 5813

QY 6303 ARAQEWTVPIR-MEGAAWPFAGCTGELLMDVHSHVHVRETTQRTTYTQOAI--THTARPPSMQ 6360  
 Db 5814 SLVTNTYTIKREAMGSGW-----SPCAKSRYYTITIEGLRAGKQYBEFR 5856  
 QY 6361 VTIEDVQAOI--GGTAQFEALIEGDPQPSVTWYKDSVQLVDSTRLSQOQEGTYSVLVR 6417  
 Db 5857 ITAENKHGSGKCEPTA--PVILPGDERKRRGYD-----VDEQKIVRKGKTVSS--- 5905  
 QY 6418 HVASKDAGYVTCIAQNTGGQVLCBAELLVLGDNPNPDSKQSHRRKLHSHFYEVKEEIGRG 6477  
 Db 5906 ---NYDNVYVDFIKWQYV-----PQVEIKHDE--VLDHYDIHEELGTG 5943  
 QY 6478 VFGFVKRVOKNGKILCAAKFIPLRSRTAQAYR-ERDILAAALSHPLVTGLDQRETRKT 6536  
 Db 5944 AFGVHRVTERATGNNFAAKFVMTPHESDKETVRKEIQTMSVLRPTLVNHLHDAFEDDNE 6003  
 QY 6537 LILILCSSEELDLRY-RKGWVTEAEVAVKVIQQLVEGLHYLHSHGVHLHDIKPSNIML 6595  
 Db 6004 MWMIYEFMSGGELFEKVADEHNKMSDEAVEYMRQCKGLCHMHENNYVHLDLKPENIMF 6063  
 QY 6596 VHPAREDIKIDGFAQNTITPAELQPSQYSGSFVSPELIQONPVSEASDIWAMGVISL 6655  
 Db 6064 TTKRSNELKIDFGLFAHLDPKQSVKVTGTGTAFAAPEVAEGKPGVGYTDMMSVGVLSVI 6123  
 QY 6656 SLTCCSPFAGESDRATLLNVLEGRVSWSSPMAAH--LSEDAKDFIKATLQRAPOAPPSAA 6713  
 Db 6124 LLSGLSPFGENDDETLNVKS--CDNMWDSAFSGISEDGKDFIRKULLADPENTMTIH 6181  
 QY 6714 QCLSHWFLK-SMPAEEAHFINTKQLK 6739  
 Db 6182 QALEHEWLPFGNAPGRDSQIPSSRYTK 6208

RESULT 10

Tl3931  
 projectin - fruit fly (Drosophila melanogaster) (fragment)  
 C:Species: Drosophila melanogaster  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000.  
 C:Accession: Tl3931  
 R:Daley, J.; Southgate, R.; Ayme-Southgate, A.  
 J. Mol. Biol. 279, 201-210, 1998  
 A>Title: Structure of the Drosophila projectin protein: isoforms and implication for pr  
 A:Reference number: Z17815; MUID:98300339; PMID:9636710  
 A:Accession: Tl3931  
 A>Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-6658 <DAL>  
 A:Cross-references: EMBL:AF047475; NID:g3337430; PID:g3337431; PIDN:AAC27550.1  
 A:Map position: 4  
 A:Note: intron positions not resolved (incomplete sequence)  
 C:Keywords: muscle

Query Match 3.4%; Score 1392; DB 2; Length 6658;  
 Best Local Similarity 18.4%; Pred. No. 2.7e-33;  
 Matches 1567; Conservative 1067; Mismatches 2994; Indels 2902; Gaps 382;  
 QY 182 YEVRASNPGLGAASAAALVVDSDAATASRPGTSTAALLAHLORREARAEAGAPASPPS 241  
 Db 2 FRVAVNRAG-----PGEPSDKTSIIAKCFVKPFVIGE----- 36  
 QY 242 TGTRCTVTGKHARLSVCVYTGPKETVWKQDQGLVTEGRRHVYVEDAQENFVLKILFC 301  
 Db 37 -GLKNVTVKGGQITRFDIKYDGEPEPAATVWKTDLNKFDCNORICLDQLERNSSITIKKS 95  
 QY 302 KQSDRGLYTCTASNLVGGQYSSVLVVRE---PAYPEFKRLQDLEVRKESATFLCEVP 357  
 Db 96 VRKDTGKLVLSNSGTTIESEAQVVLDRPLDPGGPFPE-----EIR 139  
 QY 358 QPSTEAAWFKETRLWASAKYIGIEEGTERRLTVRNVSDDDAVY-CETPEGSRTVAELA 417

Db 140 ASHIKWKWR : : : : : PDDGG--CE----- 157  
Qy 418 VQGNLLRLPKRTA--VRVGTAMFCVELAVPGVHWRNQEBVAGG-----RV 466  
Db 158 ISGVALERMDDETGRWIPAGE-----VGP-----NETSDFKGLTPNKYKFRV 201  
Qy 467 -AISAGETRLTL-TISQCCLEDVQVAFMAGDCCTSTRFCVSAPEKPELOPPVDPVVKAR 524  
Db 202 KAINKEGESEPLETFDASC-----KNPYDPP-SPPSQPVIDDY 238  
Qy 525 MESSVILSWSPPHG-ERPVTIDGYLVEKKLGTYTIRCHEAEWATPELTIVADVAEEG 583  
Db 239 DNKSVLLKWKPPSDCGRPIT--HYIVEIKDPAPMSSEVAKTD-DPNPECNVEGLKEM 295  
Qy 584 NFQVRVSALNSFGQSPYLEPFGTWHLA-----PKLAVRTPKAVQAVEGGEVTFSDLT 637  
Db 296 VYQFRVRAVNAKGPSEPSQ-PTDNLKCHKNLKPCID-RSTFKRVTIKSGRTHKWSVDVL 353  
Qy 638 VASAGE-----W-----FLDGOALKASSVYEHCDRTHTLIREVPASLHGAQLKFVANG 688  
Db 354 GEPIPELHWSMRDDIPLTNGDRKIENV-DYHTDFSITNVLKRDGSGFYTLKAENR--NG 409  
Qy 689 IE-SRIEVRAPAP-----GLTANKPPAAAAREVRLARHEEA-----QLLAELSDAQAAV 737  
Db 410 IDRETVELVLGKPSPPKGLAVSDVTASGSKLQWKXPEDVGGVPIKEYVVEKMDTATG- 468  
Qy 738 TWLKDGRTLSPG-----PKYEVOASAGRRLVLRDVARDDAGLYECVSRGGRGIAYQLSV-- 791  
Db 469 KWRVGR--SPEKEPSPDVTG-----LSLSEYMFVRVSANV 504  
Qy 792 -QGLARFLHKDMAGSCVDNAVAGGAQFE-----CETSEAHVHVHVKDGMELGHS 840  
Db 505 EEGDAEPL--TTLGVWAKDP--FDEPNKPGTPEVTVDYDQISLKWAAFNNDGGAP 557  
Qy 841 GERFLQEDVGR-----HRLVAAVTVDQEGTYSR-----VGEDSDVDFRLRVSE 885  
Db 558 IQVILIEKKNKNTKEWEKALIPDQLEATVAGLQEYGEYQFRVIANVNAKGLSPPSDASV 617  
Qy 886 PKVFAKEQLAR-----RKLQAEAGASATLSCEV-AQAQTEVTWYDKGKLS-----SS 933  
Db 618 FOIVKYKKLPKIDRSLNKLPLIRAGKPIRYDVNVNVRGEPAPVITWYQNDKELPEELPSS 677  
Qy 934 SKV-----CMEATGCTRLVLVQOQADAG-EYSCAGGQRLSHLDVKEP 978  
Db 678 SEIKNIPNTKISIEITVRKHTGYIKIIAVNEHQDEATVEVNIAPPSRPGPLDVKO- 736  
Qy 979 KWFPAKQVAHSEVOAEGANATLSCEVAQAQAEVVMYKDKKLSLKVHVHBAKGRRR 1038  
Db 737 --VTKD-----SCFKWKKPE----DGGKPI SAYQYKEKEDKKQGRW 772  
Qy 1039 LVVQOAGKTADGYSCEARGORVSFRLHITBPKMFAPKEOSVNEVOAEGASAMLSCE- 1097  
Db 773 VPL---GRTSANDTEFDYK-----LQEGHEYQFRVKAINEEGESDPLDSD 816  
Qy 1098 --VAQAQTEVTWYDKGKLSLSSSKVGMVKGCTRLVLVQAGKADAGEYSCA----- 1148  
Db 817 SIIAK-----NPYDAASRFG-----TFNIV-----DYNEHVMVKLWEAPRS 852  
Qy 1149 -GGORVSFHLHITBPKGVFAKEQSVHNEVOAEGACTTAMLSCEVAQPOQTEVTWYDKGKLS 1207  
Db 853 DGGAPISG--YIIEKKDKFS---PIWDEI-----LSTNTSVPEATV-----EGL--- 891  
Qy 1208 SSSKVRMEVKGCTRLVLVQOQKADAGEYSCAGGQRVSFQHLITB-----KAVFAKEQLV 1264  
Db 892 -----VEGNIYQFRVRAVNAKAGSDPS-----DATEPHLAKPRLNKPINDKIM- 935  
Qy 1265 HNEVTRTAGASATLSCEV-AQAQTEVTWYDKGKLSLSSSKVR-IEAAGCMRQLVQOAGQA 1323  
Db 936 -KPIKVRAGQPVKPDVDVKGEPAFSLTWFLKETELTSTGOVRLNIDYNTKLLTDDTRK 994  
Qy 1324 DAGEYTCAGG-----QRLSHFLDVSEPKAVFAKE--QLAHRKVQAEA 1364

Db 995 QSQYQKLAENINGVDEAVVEVILDKPSKPEGPLEVSD-----IHKEGCKLKWPKDDG 1050  
Qy 1365 GAIALTSCVEAQAQTEVTWYDKGKLSLSSSKVRMEAVGCTRLVLVQOQADAGEYSCA 1424  
Db 1051 G-----IPITGYIEKMDTATG-----WVPAGSDPEKYDIEI 1084  
Qy 1425 GG-----QRLSFSLDVAEPKVPFAKQPVHREVOAAGASTTILSCBEVAQAQTEVVMYKDGK 1480  
Db 1085 KGLDPNHRVQFRVKAENEE---GESEPLETSAITAKPPFDVSAPPGIPELE-DWDEHHV 1140  
Qy 1481 KLSFSKVRMEAVGCTRLVLVQOQADAGEYSCAGSQRSLFHLHVAEPKAVFAKEQPA 1540  
Db 1141 KLKWEPPIRDGSGPITNVII--EVMDKDPGEF-----VKAV-ETDSPV 1180  
Qy 1541 SREVOAEGTSATLSCEVAQAQTEVTWYDKGKLSLSSSKVRMEAVGCTRLVLVQOQAD 1600  
Db 1181 CKGV-----KKEEGQQYKF-----RVRAVNAKAGPSD 1208  
Qy 1601 AGEYSCKAGDQRLSFLHVAEPKVPVFAK---EQPAHREVQAEAGASATLSCEV-AQAQTE 1656  
Db 1209 PSEQT-----NWHVAKPR--FLKPHIDRVNLKPVIVTKTGLSISLUDINIRGEPAK 1256  
Qy 1657 VTWYDKGKLSLSS--KYRVEAVGCTRLVLVQOQADAGEYSCAGGQRLSFLHVAELE 1715  
Db 1257 VEWFFNNSSVTSDHSVKIDNVYNTKFFVMRAQRSQSGKVIKATNE---VGEDAELE 1313  
Qy 1716 POISERPCR-REPLVVKEHEDIILTATLATPSSAATVTWLK-----DQVEIRRSKRH 1765  
Db 1314 VTVLKPKPKPLQVND-----ITKHSCKLKWEPDDGGSGSDIYIEIKLDPH 1363  
Qy 1766 E-----TASOGDTHLTVHGAQVLDIAIYSCRVA---EGQFPFVQVBEVAAKFCRL 1815  
Db 1364 TQWMLPCGKSTPEAKVIGLH-----EGKAYKFRVRAVNAKEGESEDELEK- 1409  
Qy 1816 EPVCGELGOTVTLACELSPACAEVVM---RCGNTOP--RVKGRFQWVAEGPVRSLTVLGL 1870  
Db 1410 -----PIIAKNPYDEPDRPGPEPTNWDKDFVDLAWDPPKN----- 1445  
Qy 1871 RAEDAG-----EYVCSRDDHTSAQL-TVSVPRVVKFMSGLSTVVAEAGEATFOCVVSPS 1925  
Db 1446 ---DGGAPITQYVIQMRDKSGRAWDSATVPG--DKCNGTIVTG--EEGHEYEFRIV--- 1495  
Qy 1926 DVAVVWFSDGALLQSEKFAISQSGASHSLTISDLVLEDAQITVEAEGASSAALRVRE 1985  
Db 1496 -----AVNKAGSPDSPSVSKVI-----A 1514  
Qy 1986 APVLFKKLEPQTVEE---RSSVTLEVEL---TRPWPBLRWRNATALAPKNVIEHAEG 2039  
Db 1515 KPRFLKPHIDRKNLQKIMRSGQMLHIDALIKAEPPAKVTWYNTKTEIKTSDHIKIENED 1574  
Qy 2040 ABRLVLHNVGFADBGFCGCEPDDKTCAKLTVMRQVRLVRLGLOAVEARQGTATMEVQ 2099  
Db 1575 YKTTFIMPVKVRADGIVYIVTAKND-----SGSDTVEVE 1608  
Qy 2100 LSHADVDSWTRDGLRFOQGPTCHLAVRG-PMHITLISGLRPEDSG----- 2144  
Db 1609 L-----EVLCKSPKPKPLAVSNVTAETLHLKWEKPEDDGGDPIQYILVERMDT 1657  
Qy 2145 -----LMVFKAGVHTSARLVLTTELVSFSRPLQDVVT 2177  
Db 1658 ETGRWVPLVTTKTPEADVTLGTEGKEYLFRVKAENVSEGESPLEVTDIPTKAKNFDAADT 1717  
Qy 2178 TEKEKVTLECLSRPNVDVRWL---KQGVLRAGTKWMAIAQAQACRSITIVRCEPAD--- 2231  
Db 1718 PKKPIV---DWSGNHCDLKWEAPEDDG-----GA--SITGIVVERKDPT 1758  
Qy 2232 ---QVYVCDADHDAQSSASVKVQGRTYTLIYRRLVAEDAGE----- 2269  
Db 1759 GKMQKALETSTPDCKARVNDLIAGNKYQ--FRIMAVNAKAGSKSPSESDQMTAKDRFAPP 1816  
Qy 2270 -----IOFVAENAESRAQLRVK--ELPVLTVRPLRDK-----IAMEKHGVL 2309  
Db 1817 KIDRNIKIDITSKAGQHIRFDIKUSGEPPATKVM-LHNKARLENDSDNYNIDMESYRTEL 1875

QY	2310	ECQVSRASQVRWFKGSQEL-----DAGDVKTSQAQFFVEQSIIIVRGLQDVTVMEEAPAWFECET	2350
Db	1876	TVPIISK-----RPHSGKYTLKAENESGRDEASFVIVLDKPGP-----PEGPLR-----VTD	1922
QY	2351	VHADEDEDTYTC-----DAGDVKTSQAQFFVEQSIIIVRGLQDVTVMEEAPAWFECET	2402
Db	1923	VHKEG-----CKLKNMAPPDDGGLPID-HYIIEKXMDVESGRWL-----PSGRFKSFA	1969
QY	2403	SIPSVRPKWLGLKTVLQAGNGVGLQEQGTVHRLMRLRTCTMTGPVHFTVKGSRSSARL	2462
Db	1970	ELNLEPS-----HEYKFRVLAVN-----TEGES-----	1993
QY	2463	VNSDIPVILTRPLEPKTGRELQSVLVSCDF-----RP-APKAVOMVKDTP-----SPSEKF	2514
Db	1994	-----EPLTGB-----QSVIAKNPFDPGPKGTPEAVDMDKHDVLVWMPPIINDG	2038
QY	2515	KMSLEGQMAELRILRLMPADAGVYRCQASAHSSTEVTVEAREVTVTPLODABATEEGW	2574
Db	2039	GSPIITGVYVEKR-----EKGTDKWKIGTEITIPCLGECKAT-----	2075
QY	2575	ASPSCELSHEDBEVWSLNGMPLYNDSPHFHISHKGRHTLVLSIORADAGIVRASSLUKV	2634
Db	2076	-----VPTLNEC-----EYEFVRKAINAAGPEPSDASKPI	2107
QY	2635	STGARLEVRKVPVF--LKALDGLAERGTALQCEVS-DPEAHVWVRKDGVQLGPSDK	2691
Db	2108	ITKPR--KLAFTLDPTKNIITYNFKSGEPIFLINISGEPADVTWQNNKSVQITSP	2164
QY	2692	YDFLHTAGTGLVYVHDPEDAGLY--TCH--VGSEETRARVRHDLHVGITKRL--T	2744
Db	2165	SHIENLPYNTKYINNPNRKGTLGYKISAHNFYGDQVEFOINI-----ITKPKPGGP	2218
QY	2745	MEVLE--GBSCSF-----ECVLSHESASDPAMTGVGKTVGSSSRFQATRQG	2789
Db	2219	LEVSEVHKOGCKLWKVKKDDGGEVPVESLYVEKFPDPTGILWLPVGRSGPYNVVDGLVPG	2278
QY	2790	RKYILVREAPSDAGEVSVFVSRGLTSKASLIVRERPAALIKPLEDQWVAPGEDVELRCE	2849
Db	2279	HDYKERVKAVNKEGESEPLETLGSIIAKDPPFSVPTKPGV--PEPTDWTANKVELAWPEP	2335
QY	2850	LSRAGTPVHVLKDKRAIKRSQKXVVCBGTM-----AMLVIRGASLKADAGEYCEVEAS	2903
Db	2336	ASDGGSPIOGY-----IVEVKDYSLPWEKALETNSPTPTATVQG--LIEGNEYQFRYVAL	2389
QY	2904	KSTASLHVKEKANCFT-----EELNLOVEEKGTAFTVCK--TEHPAATVTR	2949
Db	2390	NKGNLSEPDSPKIFTAKKRYIAPKLDNRNLNITL--SSGTALKLDANITQCPAPKVEW-	2447
QY	2950	KGULELRASKHOPSQEGITL-----RLTISALEKADSTYTCIDICQASRAQLLVQG	3002
Db	2448	-----KLSNYHLQSGKNTIETPDYTKLVIRPPQRTDSGEYLVTANTSGKDSVLVN-	2500
QY	3003	RVVHIIED-----LEDVDV--QEGSSATFCRISPAHYEPVHVF-LDK-----TP	3044
Db	2501	---VWITDKRSPNGPLQISDVHKEGCHLWK-K-RSHDGGTPIEYFDKLEBPTGWCWIP	2556
QY	3045	LHANELNEIDA---OPQGYHVLTLRLQALKDSGTYTFEAGDQORASALRVTEKPSVFSRE	3101
Db	2557	SCRSTAPQVDVTCGLSPGNEY--KFRVSANVASGESQPLVGDBSIVARNPFPDEPKPENLK	2614
QY	3102	LTD-----ATTIEGEDTLVTCSTCDIPMCWTDKGKTLRG-----SARCQLS-	3144
Db	2615	ATDWDKHDVLAWTPPVIDGGSPF-----SCYIEKODKYGKWERALDVPADOCKATI	2667
QY	3145	---HEGHRAQLLITGATLQDSRYKCEAGGACSSSIVRVHAR-PVRFQEALKDLVLEGG	3200
Db	2668	PDLVEGQTYKFRVSANNAAGTG---EPSDSTPPIAKARNKPPIIDRSSLVEVRIKAGQ	2723
QY	3201	AATLRCVLSVAAP-VWVCYGNVLRAPBGKYSLROEGAMLELVLRNLRPQDSGRYSC---	3256
Db	2724	SFTFDCKVSGEPAPQTKLLIKKEVYSKDNVTVNDVNTYNTKLKVNASATSDSGIYTFAE	2783

Db 3574 --SEPSSEDMIRKPRYLPPKIITPLNEVRIKGLIFHTDIHFIGEPAEATWLNPNP 3631  
QY 4290 LOANEMNDITVEQGTLLHLLTKHLLKHTLEDAGTVSHVGTCSSEAKYLVKTAKNTVVRGLENV 4349  
Db 3632 LLSNDRSIT--SIGHUSVHTVNCORSDSGITHL-----LUNSSGI 3672  
QY 4350 BALEGGEALF-----ECOLSPEVAHAHTWLLDDEPVRTSENAEVVFFENGRLHLLL 4401  
Db 3673 D--EGSFLVLDPRGPEGMEYEITANSVTISWKP----- 3708  
QY 4402 KNLPRQDSCTVTFIAGDWTSAFTVR-----GWRLEILLEPLKNAAVRAGAQAQRTCTL 4455  
Db 3709 ----PKON-----GGSEISSYVIEKRDLTTHGGW----- 3733  
QY 4456 SEAVP--VGEASWYINGAAVQDDSDWTVTADGSHQALLLSAOPHHAGEVTFACRDAVAS 4514  
Db 3734 ---VPANNVYSAKYNHAVPR-----LLEGTYELRVMAENLOGRSDPLTSDQPVAK 3783  
QY 4515 ARLTVLGLPDPPEDAEVVAHSSHTVTLSWAAPMSDGGGLCGYRVVEKAGATQW-RLCH 4573  
Db 3784 SQYTVPGAPGKP---ELTSDKNHITIKWKQPIGNGSGPIIGYDIERRDVTGWIKING 3840  
QY 4574 ELVPGPECVGLAPGETYRFRVAAGVGVGAGEPVHLPQTVRLAEPKPVPPPOPSAPESR 4633  
Db 3841 QPVTAEVQDDRVTSNHQYVIRISAVNAAGN-----KTSEPSAIFNARPLEKPR 3891  
QY 4634 -----QVAGEDVSLELV--ABAGEVIWHKMERIQGGFVVSQGRQOMLV 4681  
Db 3892 FYFDGLIKRIKVRAGEPVNLNIPISGAPTIETWEKRGDLKEBGRKISYETNSERTLFR 3951  
QY 4682 IKGFTAEQGVHCHGLAOGSICPAANTFOVALSPASVDEAPOPSLPPEAAQEGDLHLLWE 4741  
Db 3952 IDDSNRDRSGKI-----TVTANFEGKDTADIEVIVDKFS-PP-----EGPLSY--- 3995  
QY 4742 ALAKRRMSRPTLDSIS---ELPEEDGRS-----QRLP---QBAEEV 4778  
Db 3996 -----TETAPDHISLHWYSFKDDGSDITGYIIEFTFEGVDDMKVPVPGTCPTNFT 4046  
QY 4779 APDLSEG-----YSTABELA-----RTGDADLSH 4802  
Db 4047 VKNLVEGKXYFRIRAEENIYGASEALEGKPVLVKSPDFPCAPSQPTISAYTPNSANLEW 4106  
QY 4803 TSSDD-----ESRAG-----TPS-----LVTYLKKAGRPCTSPASKVGA 4837  
Db 4107 HPPDDCGKPIGTIVYVERREGEWIKCNVPTENTSYTVSNLRDGAIFYFRVLAVNEAG 4166  
QY 4838 PAASVVKPQOOQBPLAA-----VRPPLGLDSTKDLGDPMDKAANKIQAFAKQYKV 4888  
Db 4167 PGHFS-KP--SDPMTAEHORYRDPPEPPKPDRIITRNGVTLNWRPRTDGKRIKGYV 4222  
QY 4889 RKENKQOGB-----PMPSHF----- 4904  
Db 4223 --EVRPKNGKDWKTVNDIPINSTVYVPSLKEGEYSFRVVAENECRSDPSKPSQPIITI 4280  
QY 4905 -----GDTEAQVGDALRLCEVVASKADVRARWLKDGVELTDRHHHIDQLG 4950  
Db 4281 BEQPNKPCWELGKVRDIVCRAGDDFSIHVPYLAFFKPNFAWYNDNMLDDNNRVH-KHLT 4339  
QY 4951 DGTCSLLIAGLDADAGCYTCQVSKFGQVTHSACVVVSGSEASESSGELDDAFRA 5010  
Db 4340 DDAASVVVKNKRDSQYRLQFKDTSG-----FDTA 4371  
QY 5011 ARRLHRLFRKSPAEV-----SDELF-----SADEGPA-----BPEEPADWQTYRED 5054  
Db 4372 TINRVLDPRPPTRLRADVPFSGSLTYNNPPNDGGSALQNYVIEKKE-ARSSTWSKV 4430  
QY 5055 EHFICIRFEALTEARQAVTRPOEMFATLIGVIELVQEGPRRVECMISKETP--APVVP 5112  
Db 4431 SSFCTVLIF---VRIRNLVLNKEYDFRVI---AENKYGQSDPANTSEPILARHFFDIPNTP 4484  
QY 5113 PEPLFSLTSDAARVFLTELQNBQVODGYPVVSFDCVVTVGPMPSVRWFKDGKLEEDDHY 5172

Db 4485 GIPHGIDSTEDSITIATWTKPKH-----DG-----GSPI-----TGYIIEK---R 4520  
QY 5173 MINEQOQGHQHLIITAVVPADMVVVRCIAENSMGVSSTKAELRVDLTSTDYDTAADATES 5232  
Db 4521 LLSDDKW-----TKAVHALCPDLSCKIPNL--IENAEYFRV-----AAVNAAGQ 4563  
QY 5233 SYFSAQGYLSSREGEQTESTDEQQLPQVVEELRLQVAPGTRLAKPOLKVKGY--PAP 5290  
Db 4564 SAYSGSSDLIFCRPPHAPKITSLSIP-----DMTVIAGD---EPRITVPYHASP 4613  
QY 5291 RLYWFKDQOPITASAHIMTGKILHLEIISVTREDSGOYAAVYISNAGMAGYASARLLV 5350  
Db 4614 TASWSLNGLEVIPGGRIFKFDNSDYASMYNKSARKDETSYITILNNKSGDTSCHVTV 4673  
QY 5351 RGFDEPEEKP--ASDVHEQLVPPRMLERFTPKVKKSSITFSVKVEGRPVPTVHWLREE 5408  
Db 4674 VDRPLPPQGLNAYDI-----TPDT-----CTLAWKTPLDGSGSPITNYVVEK 4716  
QY 5409 AERGVLWAGPTPGTYTVASSAQOHSVLVLDVGRQHQCTYTCIASNAAGQALCSASLHV-- 5466  
Db 4717 LDNSGSWV-----KISSFVNTHYDVMGLEPHYKYNFRVRAENQYG---LSDFLDIE 4766  
QY 5467 -----SGLPKVEQEKVKEALIST-----FLOG-----TT 5491  
Db 4767 PMVAKHQFTVPDEPQPKVIDWDSGNVTLIWTRPLSDGSGRIQGVQIEYRDILNDSSMNA 4826  
QY 5492 QAIASAQGLSETASFADLQORKEEPLAAKEALGHLSLABVGTTEFLQKLTQSITEVMSAKI 5551  
Db 4827 YDIYIKGTQYLYNLINGSYEFRIKAKNAAG---LSKPSPSLRFKL----- 4871  
QY 5552 TQAKLOVPGSDSDSKTPSPASRGRSRSPSSISOESSESEDGDAR--GSIPO----- 5603  
Db 4872 -KGKFTVP-----SPPGAPQVTRVGNVYDLKWEKPLRDGSGRITGVIIERRDIG 4921  
QY 5604 -IYVVTADYPLGABQDAITLRE-GQY-----VEVLDAHP 5637  
Db 4922 AVWKRCNDYVLDITETVNMNIEMGDYEFYFVAVNSAGRSPSLCTWPIKVCEVLGGKKP 4981  
QY 5638 LEWLVRTKTPSSPSRQWSPAYLDREL-----KLSP--EW--GAEAPEFGEAVSE 5687  
Db 4982 -DWITRL-----QDKVAPFGKDYTLQCAAGSKSPFTARWLKNGKIEQWGRMTCO 5031  
QY 5688 DB---YKARLSSV-----IQELLSSEQAFVEELOFQSHHLQHLERCP----- 5727  
Db 5032 SKDGVFRLHISNVQDGDYTCAMNS-LGFVNTSGYLKIGS--PIINRCPSELKLPEDG 5089  
QY 5728 --HVPPIAVAGOK--AVIFRNVDRIGRHSFLQELQOQDTHDDVAMCFIKNAQFAEOYLE 5763  
Db 5090 NSKIKIFYSGDQPLTVILKKNNEV-----ISDSNDTHV---KVNIFDDYVA 5133  
QY 5784 FIV-----GRVQAESVWVS-TAIOEFYKKAEEALLAGDPSQPPPP----- 5823  
Db 5134 IVIANIVKSDGPGYQIEFTNESGATGEFY-----VHITGMPAPTGMGSIYINKNSC 5187  
QY 5824 -----PLOHYLEQPERVQRYQALLKELRNKARNQNCALLEQAAYVASAL 5870  
Db 5188 MLNMRPPYDGLKVSHTYTERKDVSSPHWITVSTCKDTAFNVQ--GLIENQBYIFRVM 5245  
QY 5871 PQAENKLHVSLMENYPG-TLEALGEPIRQGHFTVWEGAPGARWPKGNHREHVFNLH 5929  
Db 5246 -----AVNENGMGPPEGL-NPIRAKDPIDPSPPAV-----LAKSLSEE 5284  
QY 5930 VICKP-----RRSRTDTVYVFRNMMLKSSIDLNDQVGGDDRAFE 5970  
Db 5285 TLICNPEWKKPSDGAHTQGYWIDKREVSKHMGVAVNATICAQINCINLIEGRQYEF 5344  
QY 5971 VWOEREDSVRYLLQARLAIKSSWKEICGIQORLALPVWRPDPFEEL-ADCTAELGE 6029  
Db 5345 IFAQN-----VAGLSAIVR-----FTSQNRSTGGFPASIDCTAHANGIQ--NH 5389  
QY 6030 TVKLACRVGTGPKPVISWYKDGKAVQVDPHHILIEDPDG---SCALLDSLT-----GV 6080  
Db 5390 NAOFTCTIPGVKPTISWYKGAERIS-----NGAGITCTQKTSLPKYDYVFE 5438

RECEIVED OCT 19 1966

## RESULT



Db 2241 PGPIPPVRISSSTVAEGQTLDLSCVVGQAHAQVWYKKG-----GSLPARHQVGR--S 2294  
 Qy 1038 RLIVQOAGKTAGDYSCBA-RGQVVSFLHIT-----EPKMPFAKE 1077  
 Db 2295 RLYTFQASFPADAGQVVCASNGMEASITVTGTQGANLAYPAGSTOFRIEP----- 2347  
 Qy 1078 QSVHNEVQAEAGASAMLSCEV-AQAOTEVWYKDGKLLSSSSKVMGVKGTRELVLPQA 1136  
 Db 2348 -----SSSQVAEGQTLDLNCVVPQSHAGVTHKKG-----GSLPVHQTGSLRLY--QA 2397  
 Qy 1137 GKADAGEVSCAGGQVVSFLHII---TEPK-----GVFAKEQSVHNEVQAEAGTTAMLS 1187  
 Db 2398 SPADSGEVVCRVLGSSVPLEASVLVITIEPAGSVPALGVTPTVRIESSSSQVAEGQTLDLN 2457  
 Qy 1188 CEVA-QPQTEVWYKDGKLLSSSKVMGVKGTRELVVQVQKADAGEVSC-----EAGG 1242  
 Db 2458 CLVACQAAHQVTHKRGSLFA---RHQVFGSLRL--QVTPADSGEVVCRVVGSSGT 2511  
 Qy 1243 QRVFQLHITEPKFAVFAKEQVHNEVTEA-----GASATLSCEVA-QAOTEVWYKDG 1294  
 Db 2512 QEASVLVTI-QORLSGSHSQGVAVPVRIESSSSASLANHTLDLNCVASQAPHTITWYKR 2570  
 Qy 1295 GKLLSSSSKVRIEAAGCNRLVVOQAGQADAGEVTC-----AGQRLSFHLV-----SE 1345  
 Db 2571 GGSGLPS-----RHQIVG--SRLRIPQVTPADSGEVVCHVSNAGAGSRETSLVITQSGSSH 2624  
 Qy 1346 PKAVFAKEQLAHRKVQAEAGAIATLSCEVA-QAOTEVWYKDGKLLSSSKVMGVKGTRELV 1404  
 Db 2625 VPSVSPPIRIESSSPTVEGQTLDLNCVVAQPQAIIWYKRGSLPS-----RHQTHGSH 2680  
 Qy 1405 RRLVVQACQADAGEVSCAGQRLVSLDVABPKVFA-----KEQPVHREV 1452  
 Db 2681 LRL--HQMSVADSGEYVCRANN-----NIDALEASIVISVSPSAGSPASFGSSMPIRIES 2733  
 Qy 1453 QAQ---AGASTTLSCV-AQAOTEVWYKDGKLLSSSKVMGVKGTRELVVQACQAD 1508  
 Db 2734 SSSHVAEGETDLNCVVPQQAHAQVTHKRGSLPSHQTR-----GSLRLHLHVS PAD 2787  
 Qy 1509 AGEVSC-----BAGSQRSLFHLHVAEP-KAVFAKEQPASREVQAEAGT 1550  
 Db 2788 SGYVCRVWGSGLPEASVLVTEASG---SSAVHPAPGAPPIRIEPPSSRV-AB-GQ 2842  
 Qy 1551 SATLSCEV-AQAOTEVWYKDGKLLSSSKVMGVKGTRELVVQAGQADAGEVSK-- 1607  
 Db 2843 TLDLKCVPQQAHAQVTHKRGSLPARHQVH-----GFLRL--NQVSPADSGEYSQV 2896  
 Qy 1608 --AGDQRLSFHLV--AEPKVFAK--EQPAHREVQAE--AGASATLSCEV-AQAOTEV 1657  
 Db 2897 GSSGTLASVLVTEPSSPGPIAPGLAQPIYIEASSSHVTEGQTLDLNCVVPQQAHAQV 2956  
 Qy 1658 TWYKDGKLLSSSKVMGVKGTRELVVQAGQADAGEVSCAGG-----QRLSFLHVA 1712  
 Db 2957 TWYKRGSLPA---RHQTHGSLRL--HLVSPADSGEYVCRASGFGPEQASFTVTP 3010  
 Qy 1713 ELE-----POISERPCREPLVWKEHEDILTLATLTPSA-ATVTM-----LKDG 1757  
 Db 3011 PSEGSYLRSPVISIDP--PSSTVQGGQDASFCKLIHGAAPISLEWTRNQELEDNV 3067  
 Qy 1758 EI-----RRSKRHET-----ASCG-----DHTLTVHGAQVLDLSAISCRVGAEG 1797  
 Db 3068 HISPNGSIITVTGTRPSNHGTYRCVNASNAYGVAQSVNLSVHGPPTVS-----VLPEG 3120  
 Qy 1798 QDFPVQVEEVAKFCRLLEPVCGLGTGTVLACELSPACAEVWRCNTOPRVGKRFQMV 1857  
 Db 3121 -----FWVKVGVKAVTLEC-----VSAGEPRSSARWRI 3149  
 Qy 1858 AEGPVR-----SLTVGLRAEDAGEVYVCSRDDHTSAQVTSV-----P 1896  
 Db 3150 SSTPAKLEQRTYGLMDSHAVLIQISSAKPSDAGTVCLAQNALGTAQVQVEIVDTGMAP 3209  
 Qy 1897 RVVKFMSGLSVVAAEGGEATFCQVSPSPDVAVV-WFRDQALLOPSKFAISQSGASHSL 1955

Db 3210 GAPOVQAEBAELTVEAGHTATLRCSATGSPAPTIHWSKLSPLPWQHRL-----EGDTL 3263  
 Qy 1956 TISDLVLEAGQITVBAEGASSAALRVREAPVLFKKKLEP--QTVEERSSV---TLEV 2009  
 Db 3264 IIPRAQDSGGQYICNATSPAGHA-----EATILHVESPPYATTVPHEASVQAGETVOL 3318  
 Qy 2010 E-LTREWPEL--RWTRNATAL---APGKNVEIHAEGARHRLVLHNVGFPADRGFFGCETPD 2063  
 Db 3319 QCLAHGTPTLTQWSRVGSSLPORATARNELIHFERAPE-----DSGRYRCRVTN 3369  
 Qy 2064 DKTOAKLTIVEMRQVRLVRG---LQAVEAREOGTATMEV--QLSHADVDSG----- 2108  
 Db 3370 KVSABEAFQAL---LVQGPGLSPATSPAGSTPTVQTPQLETKSIASVEFHCAVPS 3425  
 Qy 2109 -----WTRDGLRFOGQPTCHLVRGPMHTLTSLRPEDSGLVVFAEG---VHTSA 2157  
 Db 3426 DQOTQLRWPKEG---GQLPPGHSVQDG---VLRIONLDOSCOGTTCQAHGFWGKAQASA 3479  
 Qy 2158 RLVTTELP---VSFSRPLQDVVTTTEKEKVTLEC-ELSRNVDVVRMLKOGVELRAGKTMAI 2213  
 Db 3480 QLVIAQALPSVLINIRTSVQVVV--GHAVEFECLALGDPKQVTVSKVGHRLRP--IV 3534  
 Qy 2214 AAGACRSITIVRCFADQGVVCDHDAQSSASVKVQGRTYTLIYRRVLASDAGEIQV 2273  
 Db 3535 QSGGVVR---IAHVELADAGQYRCATNA----- 3560  
 Qy 2274 AENAESRAQLRVKELPVTLVRLDRKIAMKHEKRGVLECOVSRASA-----OV 2320  
 Db 3561 AGTQSHVLLVQALP-----QISMPQ-----EVEVPAGSAVFPICIASGYTPDI 3606  
 Qy 2321 RHFKSGQELQPGKVELVSDGLYRKLITSDVHAEDDTYTCD---GDVKTSAQFFVBE 2376  
 Db 3607 SWSKLDGSLPPDSRLR-----NNMLPSPVPQDAGTYVCTATNRQGVKAFALQVPE 3660  
 Qy 2377 QSI 2379  
 Db 3661 RVV 3663

RESULT 12

S20901

titin - rabbit (fragment)

C;Species: Oryctolagus cuniculus (domestic rabbit)

C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 18-Jun-1999

C;Accession: S20901; I46520

R;Label: S.; Gautel, M.; Lakey, A.; Trinick, J.

EMBO J. 11, 1711-1716, 1992

A;Title: Towards a molecular understanding of titin.

A;Reference number: S20897; MUID:92258380; PMID:1582406

A;Accession: S20901

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: mRNA

A;Residues: 1-6805 <LAB>

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992

R;Label: S.; Barlow, D.P.; Gautel, M.; Gibson, T.; Holt, J.; Hsieh, C.L.; Francke, U.;

Nature 345, 273-276, 1990

A;Title: A regular pattern of two types of 100-residue motif in the sequence of titin.

A;Reference number: I46520; MUID:90238553; PMID:2129545

A;Accession: I46520

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 4235-5250 <LA2>

A;Cross-references: EMBL:X17329; NID:gi1756; PIDN:CAA35207.1; PID:g930251

C;Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pr

C;Keywords: muscle

Query Match

Best Local Similarity 2.5%; Score 1046.5; DB 2; Length 6805;

Matches 1394; Conservative 968; Mismatches 2748; Indels 2411; Gaps 346;

Qy 17 KAFVSVVGKDATLSQIVGNPTTPQVSW-----EKDQPVTAGAFRLAQDGLRL 67

Db	476	KLVVVRAGCIRLPAIVRGRPAKVTWRKVGIDNVVRKQGVLDVDTMAF-----L	525
Qy	68	TILDALGDSGOVYCRARNAIGFAAAGLOV-----DAEAAAC-----	105
Db	526	VIPNSTRDSDSGKSLTLVNPAGEKAVFVNVRLDTPGVPDGLKVDVTKTSCHVSWAPPE	585
Qy	106	---AEQAPHL-----LPTSIYRE-----GSEATERCRV-----	133
Db	586	NDGGSQVTHVIVEKRDABRTWSIVNPEVFKTCQVNLVPGNEYFRVTVAVNEYPGVP	645
Qy	134	GGSPRAVSVKSGORRLCEPDGPR--VRVEELGEASALIRIAARPRDGGT-----	186
Db	646	ADVPKPVJ--ASDG--LSEPDPPKLEVTWTKNSATLAWLPRLDGGAKIDGYIISYRE	701
Qy	187	ENPIGAASAAAALVVDSDAADTASRPQ-----TSTAALLAHLQRRREARAGAPASP	239
Db	702	EDQPADRWTEYSVVVKDLSLVLTKLKGKGYKFRVAARNAVGVSLPREAGVYBAKEQLIP	761
Qy	240	PS-TGTRCTVTEGKHARLSYVTGEPKPTWKK-DGOLVTEGRHVVVEDAQENFLK	297
Db	762	PKILMPQIITIKAGKLIIEAHVVGKPOPICKWKKGEDDVTS--HLAVHKAESSILI	819
Qy	298	ILFCQSDRGLYTCTANLVGQTVSSVLVVRE-----PAVPFKRLQDL-----	342
Db	820	IKDVTBKSDGYSLTAENSSGTDQIKIVMDRPGPPQPPFD--ISDIDADACSLSMHI	877
Qy	343	-----EVREKESATFLCEVPO-----PSTEAAW-FKEETR-----	371
Db	878	PLEDGGSNITNYIVEKCDVSRGDWVTALASVTIKSICIKLIPQEVFVRVRAENRFGIS	937
Qy	372	-----LWASAKYIGIEEGTERRLTVRNVSAADDVYIC-----ETPEGSRTVAELAVQ	420
Db	938	EPLQSPKMLAQFPGVPSEPKNARVT---KVNKDCIFVAMDPRDSDGGS-----	987
Qy	421	NLLRKLPRKTA--VRVGDTFMFCVELAVPGVPHLNRQEVVAVGGRVALSABTRHILT	478
Db	988	YLIERKGRNLLWYKANDTA-----VRSTEYPGAG-----LVGGLSYSPR	1027
Qy	479	ISQCCLEVDGVAFMAGDCQSTRFVCSAPRKP-----PLQPPVDPVVKARMSSVI	530
Db	1028	I--YALNKGSS-----SPSKPTEYVTAETPVDPPGKPEVIDVTKSTVS	1069
Qy	531	LSWSPPH--GERPVTIDGVLVEKKLGTVTWIRCHEAEWATP-----ELTVADVAEE	582
Db	1070	LIMARPKHDGSK---IIGYFVEACKLPDQKWRCN-----TTPHOIPEEYVTGLEEN	1121
Qy	583	GNFQFRVYSALN--SFGQSPYLERPGTVH--LAPKLAVRTPLKAVQAVEGGEVTFSDVLT	637
Db	1122	AOYQFRAIAKTAVNISQPSSELTPTVTHAENVPRIDLVSAMKSLLTVRKAG--TNVCLDAT	1180
Qy	638	V-----ASAGEWFLDQALKASSVYIEHCDRTRHTLTIREVPASLHGAKLFVANGIES--	691
Db	1181	VFGKPMPTVSWKSGTVLKPAGKIKMAQNRNLCTLELFSVNRKDSGDYTTIAENSSGSKS	1240
Qy	692	SIRMEVRAAPGLTANKPPAAAAREVLARLHEEAQLLA---ELSDQAAVT--MLKDGRTL	746
Db	1241	ATIKLVLDREPG-----PPASVK---INKMYSRMLSWPPELDDGGSEITNIVDKRET	1292
Qy	747	SPGPKYEQA-----SAGRRVLL-----VRDVARDAGLYECVSRGGRJAYQLSVOGLA	795
Db	1293	SSRNAQVSANVPITSCSVEKLIBGHEYQFRAICENKYGVGDPVFTBPAIAKN-----	1345
Qy	796	RFLKDNAGSCVDVAGGPAQFECETSEAHVHVHVKDGMELHSGERFLQEDVGRHRL	855
Db	1346	---PYDPPRCDDPPVIS-----NVTKDHMTWSWKPPADDDGSGSPITGYLLEKRET--HAV	1394
Qy	856	VAAVTVRQD--EGTYSRVRGSDSDVDFRLVS-----EPKVVFAKEQJ-----	895
Db	1395	NWTKVRNKPVIERTIKATGLQEGTEYFRVTAINKAGGPKPSDASKAVYAQDPLYPGPP	1454
Qy	896	ARRKQBAAGASATLS-----CEVAQAQTEVTWKDQ--KLLSSSKVC	937
Db	1455	APFKYDTTSSVSLSKGPKPAYDGGSPITGYLVEVKGADTD--NWRNCLPKL---OKTR	1510

Qy	938	MEATGCT-----RRLVVQQAQADAGEYSCEAGGQRLSFHLVDVKEPKVVFQAKQVAHS	990
Db	1511	FEVTGLAMTEYQFRVYAVNVKGYSDPSDVP-----DKHCPKOILLIPPE---G	1555
Qy	991	EVQAE-----AGANATLSCEV-AQAQAEVMMYKDGKLSLSSKLVHVEAKGCRRLVV	1041
Db	1556	ELDADLRKTLILRAGVTWRLVVPVKGRPPPKITWSKENVNLREIRIGLDIKSTDFDTLRC	1615
Qy	1042	QOAKTDAGDY-----SCEARGQVSFRHLIT--EPQMFPAKEQS-----	1079
Db	1616	ENVNKYDAGKVIILTLENSCGKGYTIIVKVLDTGPPVNVTVKEISRDSAYITWDPPIVD	1675
Qy	1080	-----VHNEVOAEAGASAMLSCEVAQAQTEVTWKDGK-----KLSSSSKVGMKVGC	1127
Db	1676	GGSPINNVVVEKRDARKSWSTVTECKTSFRVSNLEBCKSYFFRYVAENEYIGIDPGE	1735
Qy	1128	TRRLVLPQAGKAD-----AGEYSCEA-----GQQRVSFHL--HITEPKG	1164
Db	1736	TR-----DAVKASETPGPVVDLKVLTVKSSCNIGWKPRSDGSRITGVYDFLTE--	1787
Qy	1165	VPAKEQS VHNEVQAEAGTAMLSCEVAQPOTEVTWKDGKLS--SSSKVRMEVKGCTRR	1222
Db	1788	-----ENKQVRMKSLSLOYSTKDLNEGKQYTFR	1816
Qy	1223	LWVQ-QVGKADAGEYSCEAGQQRVSFQHLITE--PKAVF-AKE-----QJVNHEVTRTEAG	1273
Db	1817	VSAENENGEGTPEITVVAKDDVVAPDLDLKDLPCYLAKENSFRLLKIPHQ-----	1869
Qy	1274	ASATLSCEVAQAQTEVTWKDGKLSKSSKVRIBAGCMQVLVQQAQADAGEYTCAG	1333
Db	1870	-----GKPAFSPVTKWKGEDPLATDTRVSVSSAVNTTLVVDYDCKDAGKAIT---	1916
Qy	1334	QORLSFHLDBEPKAVFAKEQLAHRKQVABAGATATLSCEVAQAQTEVTWK-----	1387
Db	1917	-----ITLKNVAGTKEGTSLIKVGKCPG-IPTPIKDFEVTABEATILKMGPPKDDG	1966
Qy	1388	KK-----LSSSKVRMEAVCGTTRLLVVQQAQADAGEYSCEAGGQRLSFLDVAEPKV	1441
Db	1967	GSEITNYILEKSDSVNNKWITCA-----SAVQKTTFRVTRLHEGMEYTFRVS-AENKYG	2019
Qy	1442	FA--KEOPV-----HREVO-----AQAGASTTSLSCVA-QAQ	1470
Db	2020	VGEGKSEPIVAKHPFVDPDAPPNIVDVHDSVSLTWDPRTKGTGSPITGYHIEFKER	2079
Qy	1471	TEVMYKDGKLSFSSKVRMEAVCGTTRLLVVQQAQADAGEYSCEAGQRLSFLHVAEP	1530
Db	2080	NSLLWKRANK-----TPIRMEDFKVT-----GLTEGLEVEFRVMAINLA--GVGKP	2123
Qy	1531	K-----AVFAKQBPASREYQAEAGTSATLSCEVAQAQTEVTW-----YKDGKLS--	1576
Db	2124	SLPSEPVVLDPIODPPGKPEVINTRNSVTL-----INTPKYDGHKLTGYIV	2172
Qy	1577	-----SSKVRMEA-----VGCTRLVVQQAQADAGEYSCKAGDQRLSFHLHVAEPKVVF	1626
Db	2173	EKRDLPSKTMKANHINVPDCA--FTVTDLVEGGKYEFRIKAKNTAGTAISAISESTGTII	2230
Qy	1627	AKEQ-----PAHRE-VQAEAGASATLSC--EVAQAQTEVTWKDGKLSSSSKVRV	1674
Db	2231	CKDEYEARTIVLDPTIKDGLTIRKAGTIVLNAISILGKPLPKSSWSKAGKDIRSDITQI	2290
Qy	1675	EAVGCTRLVVQQAQADAGEYSCEA-----GQRLSFLHVAELEPQISERPCREPLW	1730
Db	2291	TSTPTSSMLTVKYSARKDAGEYTTITATNPFGTKBEHVRVTVLDV-----PGPPGPIEI	2343
Qy	1731	KEHEDIILTATLAPSAATVW-----LKDGVEIRS-----KRHETAS-----QGDTH	1774
Db	2344	SN-----VSAEKAATLTPTPLEDGGSPISKVLEKKEKTSRLLWTVVAEDIQCRH	2393
Qy	1775	TLTVHGAQWDSATYSYCRVGA-----EGQDFPVQVEEVAAKFCRLLEP-----	1823
Db	2394	VV-----KLIQNEVLFRVSAVNHVKGEPVQSEPV--KMVDRFGPPGPKPEVSNVTK	2447



QY	1824	GTVTLACE--LSPACAEVWRCNGTQPRVGRFQWVAEGPVRSI--TVLGLR-----	1871	Db	3287	AYAEFR-----VIAKNAAGAISSPSEPSDAITCRDDVEAPRILVDVRFKDVILK-	3336
Db	2448	NTATVSMKRPYDDGSEITGYVERREBKGLRWVRAITKTPVSDLRCKVTGLQCNVYEFRR	2507	QY	2833	LEDQWVAPGEDVELRCELSRACTP-VHMLKORAKIRKQYDVVCEGTWAMVLIRGASUK	2891
QY	1872	-----AEDAGEY-----YCESRDHHTSAQLTVSVPRVVKFMSGSLSTVVAEBEGGE	1915	Db	3337	-----AGEAKLEADYSGRPPPTWMTWKDGELEGTKLEIKIADFTSTYLINDSSRR	3389
Db	2508	VSAENRAGIGPPSDASNVLMKDVAYAPGPPSNARVDTTKKSASLAWGKPHY--DGGL	2564	QY	2892	DAGEYTCVEASKSTASLHVEEKANCFTEELTNLOV-----EEKGTAVFTCKTCHPAAT	2945
QY	1916	ATFOCVVSPSVA-VWFRD--CAL-----LOPSEKF-----AISQSGASHSLTI	1957	Db	3390	DSGAYIL-----TAT-----DPGFAKHIFNVKLDVRPGPEPLAVSEVTSB--KCV	3435
Db	2565	EITGVVVEHQVGBDETWKDITGTPALRIITEFVPPDLHTKEKYNFRISAINDAGGEPAVI	2624	QY	2946	VTRKGLLELRASGRHQPSQEGTLRLTISALEKADSDTYTCDIQCAQSRQALLVQGRRV	3005
QY	1958	SDLVLEDAGQITVEAEGASS--SAALRVREAPVLFKKLEPQTVBERSSVTLVELT--	2012	Db	3436	LSWLPLDGGAKIEHYIVQKRETSRL-----AWTNVASEVO-----	3472
Db	2625	PDVEI-----VERENAPDFELDAELR-----RLTVVRAGLSIRIFVPIK	2663	QY	3006	HIIEDEDVDVQEGSSATFRCRI-----SPANYEPVHWFLDKTFLHANELNEIDAQPG	3058
QY	2013	RPNPELRWTRNATALAPGKQVE-----IHAEGARHR-----VLHNVGFRADRGFFG	2058	Db	3473	--VTKLKVTLLKNGEYIFRVNAVKNYGVGEPLSEPV-----LAVNPGPPD-PPK	3521
Db	2664	GRPAPEVTWKDDINLTRANIENTESFTLLIIEPCNRYDTGKFVMTIENPAGKSGFVN	2723	QY	3059	GYYHLTLRQALKDSGTIYF-----EAG-----DORASAALRVTE--KPSVFSRELT	3103
QY	2059	CET-----PDKTOAKLVEMRQVELVRG-----LOAVEAREOGTATMEVO	2099	Db	3522	NEVVTI-----TKDSVWVCWGHDPDSGSGSEIINYIVERDKAGRWKCNKKTVDLRFK	3577
Db	2724	VRVLDTPGPVNLNPTDIKDSVTLHW-DLPLIDGGSRIINYIVEKREATRKSYSVTWTK	2782	QY	3104	DATITEGEDLTL-----VCETSTODIPMC	3127
QY	2100	LSHADVGSWTRDGLRFQOGPTCHLAVRGPMTLTLGLRPEDSGLMVFKABGVHTSARL	2159	Db	3578	VSLGTEGHEYEFRIANAAGISAPSRTPFYKACDAVFKPGPPGNPRVLDTSRSISIA	3637
Db	2783	-----CHKC-----TKVDTGLSEGCYEFRRVMAENEYIGIGEP	2814	QY	3128	WTK-----DGKTLRG-SARCOLSHEGH-----RAOLLITGAT--LODSGRVK-----CE	3168
QY	2160	VVTELPVSFSPRLQDVVTEKEKVTLECELSRPNVDVRLK--DGVELRAGKTMALAAQ	2216	Db	3638	WKPPIYDGGSEITGVNVEIALPEEBEWKIVTPPAGLKATSYITNLVENQEKIRIYAMN	3697
Db	2815	SETKEPVKASE-----APSPDPSUNI-MDITKSTVSLAWPKPHDGGSKITGVVIEAQRK	2868	QY	3169	AGGACSSSTV-----RVHARPVRFQBAKDLVLEGGAAATLRCVLSSVAAP--VKW	3217
QY	2217	GACRSLTIYRCEPADQGVVCDADHAQASASVKQGETYTLIYRVRVLAEDAGBIQFVAEN	2276	Db	3698	SEGLGEPALVPGTAKAEDRMPLPEIELDADLKLWVIR-ACCTLRFVPIKGRPOPEVKW	3756
Db	2869	G-----SDQWTHITTVKLECVNLTGEBEYT-----FQVMVAVN	2903	QY	3218	C--YGNVNLPGDKYSLRQEGAWLELVNRLPQDSGRYSCSFGD-----	3260
QY	2277	AEBSRAQLR-----VKE--LPVTLVRLPRDKIAMXHRGVLECOV--SRASAQVWFVX	2324	Db	3757	TREHGESL-----DKASIBESTSYTLLIYGNVNRFDGSKYILTVENSSSGSKSAFVNVRULD	3812
Db	2904	SAGESAPRESRPVTKGQWLPEDLRLGIYQKLVIAKAGDNIKVEIPVLGRPPTVTWTK	2963	QY	3261	-----QTTSATLTVTALPAQFIG--KLRN-----KEATGATATLRCESKTAP	3302
QY	2325	GSQELQPKPYELVDGLYKLIISDVHAEDEDTYTCDAGDVKTSAQFFVEEQSITIVRG	2384	Db	3813	TPGPPQDLKVKVETSVTLTWDPDLLDGGSKIKNVIVEKRESTRKAYSTVATNCHKTS-	3871
Db	2964	GDQVLKOTQRVNVNTATSTILNISECVSDSGPYPLTAKNI-----VGEVGDVITIQ	3016	QY	3303	VWRKQSEITLRDGRCLYRQDQAMCELRIGLWADAAEYSC-VCEBERTSASLIRPMP	3361
QY	2385	LODVTVMEPAPAWFECETSPSV-----RPPKLLLGKTVLQAGNVGLEQGTV-----HRL	2436	Db	3872	--WK--VDQLQBGSSYFRR-----VLAENEYIGLPAETAESVKASERPULP	3913
Db	3017	VHDIPGPGPTGPIKD-EVSSDFVTFWSEPP-----ENDGGVPISNYVI	3058	QY	3362	AHFIGRLRHOESIEGATATLRCELSKAAPVWRKRESLRDGRHSLRDQDGAVCBLQICG	3421
QY	2437	MLRETCSTMTGPHVHTVCKSRSSARLVVSDIPVLTRELPKTCGRELQSVVLSCDPRAP	2496	Db	3914	P--GKI-----TLVDVTRNSVLSKEPE--HDGGSRL--GYIVEMQSKG	3953
Db	3059	EMROTDST-----TWVELATTVIRTYKA--	3082	QY	3422	LAVADAGEYSCVCGBERTSATLIT--VKALPAKFTGLRNEEAVEGATAMLCWELSK--VA	3477
QY	2497	KAVQWYKDDTPLSPSEFKFMSLEGQMAELRILRLMPADAGYRCQAGSAHSSTEVTEAR	2556	Db	3954	-----SDKWATCATVKVTEATITGLTQGEYSFRVSAQNEKGISDPQISVPVIADLVI	4008
Db	3083	-----TRLTGVVEYQPRVKAQ-----NRYGVGPGITSASI-VANY	3116	QY	3478	PVEWRKGPENLRDGRYILRQSGTRCELRQICGLAMADAGEYLCVCGQERTSATLIRALP	3537
QY	2557	EVTVTGLODAEATEEGWASFCSELSHEDBEVW-----SLNGMPLYNDSFHEISHKGRRH	2612	Db	4009	PPAFK-----LLENFTVLAGEDLKDIDVPFIG-----RPTFTVT-----	4042
Db	3117	PFKYVPGPGTPOVT-----AVTKDSMTISWHEPLSDGGSPILG--YHVERK-ERN	3163	QY	3538	ARFIEDVKNQAEAGATAVLOCELNSAAPVWRKSETLRDGRYSLRQDQGTCKELQIRG	3597
QY	2613	TLVLKSTQRA--DAGIVRASSLKVSTGSRLEVRV-----KPVVFLKALDLSA	2658	Db	4043	-----MHKDDVPLKQTRVNAESTENSSLLSIKE	4071
Db	3164	GILWQTVSKALVPNGNI FKSSGL--TDGIAYEFRRVIAENMACKSKPSKPSFVLALDIPD	3221	QY	3598	LAMADTGEYSCVCGQERTSATLITVREALPKFTFEGLRNEEAAGATAVLRCELSKQAVEM	3657
QY	2659	BERGTALQCEVSPDEAHVWRKQVQGPSDKYDFLHTAGTGLVVDHVSPEADAGLYTC	2718	Db	4072	ACREDVGHY-----VVKLSNSAG-----EATETLNAII--LDKFGPP--	4106
Db	3222	PGK---PIPLNITRHTVTLKWAQ-----PEYTG-----	3246	QY	3658	WKGHETLRDGRHSLRQDQAGCELRIGLVAEDAGEYLCMCCKERTSAMLTVRAMPSKPI	3717
QY	2719	HVGSEETEARVRVHDHVGITKRLKTMVLEVBESCSFECVLSHESASDPAMWTVGGKTVG	2778	Db	4107	-----TGPVKMD-----EVTADSIITISWEPKVD	4130
Db	3247	--GFKITSYIYVEKEDLPNGRWLKNFNSNILEN-----FTVSGLTED	3286	QY	3718	EG--LRNEEATGDDTATLWCELSKAAPVWRKRGHETLRDGRHSLRQDGRCELRIGLUA	3775
QY	2779	SSSRFOATROGRKYILVVREAA-----PSDAGEYVVSVRGLTJSKASLI-VREPPAAIIRK	2832				

Db 4131 GSSINNYVEKRTSTTTWQIVSATVARTTIKASRLKTG-----CBYQFR-IA 4178  
Qy 3776 VVDAGEYSCVCGQBERTSATLTVR-----ALPARFIEDYK-NQAREGATAVL--Q 3822  
Db 4179 AENYGRKSTVLSNBPVIAQYPFKVPVGGPTGPFVTLSSRDSMEVQWNEPVNDGGRVIGYH 4238  
Qy 3823 CELSKAAPVWRKSEITLGGDRYSR--QDGTRELOIHLGSLVADTGEYSCVCGBERTS 3880  
Db 4239 LERKERSILWVKNKTPVQTKFTTGLEEGIEYFRVSAENIVGIGKPS-----KVS 4292  
Qy 3881 ATLVTRAPQPVFRFPLQSLQAEAGSTATLOCELSEPTATVWKGGLQQLQANGRRPRPRLQ 3940  
Db 4293 ECVYARDP-----CDPPRPERIIVTRNSVTLQ---WKKPTVD 4327  
Qy 3941 GCTAELVQLQLOREDTGEYTCGSOQATSATLVTAAPVFLRELQHQVDEGGTAHLCC 4000  
Db 4328 G-----GSKITGYV-----VEKKELPDG-----4345  
Qy 4001 ELSRAGASVWRKGLQLFPCKAYO---MVQDGAABELLVRGVEQSDAGDYTCDTGHTOS 4057  
Db 4346 -----RWKASFTNIMDTQEVTLVED---HRYEPRVIAARNAAGVSEPSSESTGA 4393  
Qy 4058 M-ASLSVRVRP-----PKFKTRLOSLEQETGDIARLCCQLSDAESA-----VWOMLKEGVE 4108  
Db 4394 ITARDEIDPPRISMDPYK---DTIVVHAGESPRI-----DADIYKPIPTTQWIKGDQ8 4445  
Qy 4109 LHAGPKYEMBSOGATRELLIHOLEAKDTGEVAC---VTGGQKTAASLRVTE---PEVT 4160  
Db 4446 LSNARLEIKSTDPATSLSVKDAFRVDSGNVYLKQAVNAGERSVTVNVKLDVRGPPPEP 4505  
Qy 4161 IVRGLVDAEVTADDEBFCEVSRAGATGVQWCLQGLPLQSLQNEVTEVAVRDGRIHTRLK 4220  
Db 4506 IV-----IS 4509  
Qy 4221 GVTPEADAGTVSFHLGNHASSAQLTVRAPEVTILEPQDVQLSQQDASFCQRLSRASGOE 4280  
Db 4510 GVTAEKC-TLAWK-----PPLQD-----4526  
Qy 4281 ARWALGGVPLQANEMDITVEQGTLLHLLKHTLEADAGTVSFHVGTCSEALQKVTAKN 4340  
Db 4527 -----GG---SDIINYIVERRETSRLVW---TVVDA-----NVQTL8---CKYT---4561  
Qy 4341 TVVRGLENVALGEALFECQL-----SQPEVAHWTWLLDDP-----VRTSE 4384  
Db 4562 -----KLEGEYIFRMAVKNYGVGEPLSEPEVIAKNFVVPDAPFAPECTVTVK 4612  
Qy 4385 NAEVVFENGRLHLLKLNLRPDQSCRVTFLAGDMWTSAPLTVRGNRLLEILPLKNAVR 4444  
Db 4613 DSMVWVE-----RP-----ASDGGSEILGY---VLEKRDKEGIR 4644  
Qy 4445 AGAQRFTCTLSEAVPVGEASWYINGAAVQDDSDMTVTADGSHQALLRSAPPHHAGEV 4504  
Db 4645 -----WTRCHKRLIGELRLRVTG-LIENHNVEFRVSAENA-----AGLSEPPSPSAY 4690  
Qy 4505 TFACRDVAVASARLVGLPDPPEAEVVAHSHHTVLSNAAPMSDGGGCGYRVVVEK8 4564  
Db 4691 QKAC-DPIYK-----PGPPNNPKVNDITRSVFLWSKPIYDGGCEIQGIVIBKCDV 4741  
Qy 4565 ATGQWRILCHEL--VPGPECVVDGLAPGETYFRVAAVGVPGAGEPVHLPQTVALAEPKXP 4622  
Db 4742 SVGEWTCTPTTGINKNTNIEVEKLEKHEYNFRICAVNKAGVGDHADVPGPVIVE8---4797  
Qy 4623 VPPGPSAPESRQVAAGEDVSLIE8---VVA8AG-----FVIHKGW8RI 4663  
Db 4798 ---KLEAP-----DIIDLUELRLKIINIRAGGSLRLFVPIKGRPTPEVRKGVGDGEI 4845  
Qy 4664 QPGRFRFVWSQGRQOMLIVIKGFTAEDQGEYHCHGLAOGSICPA8-AFQVALSPASVDEAP 4722  
Db 4846 RDAAIIDSTSSFT8--LVLDNVNRYDSKGYTLTLENSSGSKSAFVTVRVLDT8---SP 4898  
Qy 4723 QPSLPPPA8Q8DLHLWEA-----LARKKRSREP-----TLDS 4757  
Db 4899 PVNLKVTEITKDSV8ITW8EP8LLDGG8KIKNYI8VEK8D8STRK8YAA8VTVNCHK88WKID8 4958

Qy 4758 ISE-----LPED8GR8QRLP-----QB8E8V8P8DL8EGY8T8A8BL8ART8GD8AD8LSHT 4803  
Db 4959 LOEG8SY8FRV8TA8NE8Y8IG8L8PART8D8P8IKV8AE8V8QP8---PGKITV8DV8TR8-NSV8LSWT 5014  
Qy 4804 SSD8ESR8AGT8SL8VTL8---KK8GR8PT8SPL8AK8P8SV8K8P8Q8Q8E8PL8A8VR8PL8GD 4861  
Db 5015 KPE8D8---GG8KIT8Q8Y8V8E8M8Q8K8E8K8W8E8C8AR8V8K8LE8AV8IT8NL8T8Q8EE8YL8FRV8-----5065  
Qy 4862 L8TK8DL8GP8SMD8KA8V8IK8QA8AF8K8Y8K8VR8K8EM8Q8Q8E8G8PM8F8SHT8FGD8T8AQ8V8GD8AL8R8ECV8 4921  
Db 5066 V8V8NE8K8GR8SD8R8SL8AP8IVA8--K8DL8V8EP8DV8K8---D8AF8S8---SYSV8Q8V8Q8DL8K8IE8VP8I 5115  
Qy 4922 ASK8AD8VR8AR8WL8K8DG8VEL8TD8GR8HH8ID8QL8G8GT8CS8LL8I8AG8LD8RAD8AG8CY8T8Q8V8NK8F8Q8VT 4981  
Db 5116 S8GR8PK8PT8IT8W8TK8O8GL8PL8K8QT8TR8IN8V8AD8SL8DT8-TLS8IK8ETH8K8DS8GH8Y8GIT8VAN8VG8QKT 5174  
Qy 4982 HS8ACV8-----V8SG8SE8E8A8E8S-----SS8G8EL8-----DD 5005  
Db 5175 AS8IE8IT8LD8K8PD8PP8K8PV8K8F8E8V8S8A8E8S8IT8LS8WN8P8LV8T8GG8C8Q8IT8Y8V8H8K8RD8TT8TV8DV 5234  
Qy 5006 AF8R8A8RL8HL8R8L8F8TK8SP8E8---VS8DE8EL8F8---LS8A8DE8P8-----AS8PE8P8-----5045  
Db 5235 VS8AT8V8ART8T8L8K8V8TK8T8G8TE8Y8Q8F8IF8PEN8RY8G8S8F8AL8D8SE8P8IV8AQ8Y8PY8K8F8G8PT8F8VT 5294  
Qy 5046 ---AD8Q8TV8RE8DE8HF8IC8IR8FE8AL8T8AR8Q8V8-----TR8Q8E8F8AT8LG8I8G 5085  
Db 5295 AT8SK8DS8MV8Q8HE8PIN8NG8S8P8IL8GY8HL8ER8K8ERN8S8IL8AT8K8V8DS8II8HD8T8Q8K8AL8NE8S8IE 5354  
Qy 5086 VE8IK8L8-----V8QG8RR8V8EM8C8IS8KET8P8AP8VP8PE8PL8S8LL8T8SD8A8P8VL8TEL8Q8N8---5135  
Db 5355 YE8FR8YA8EN8IV8GV8K8ASK8N8SEC8Y8V8ARD8---PC8DP8PG8-----T8PE8I8V8K8R8E8IT8LOW8TK 5405  
Qy 5136 EV8Q8GY8PV8S8F8CV8T8G8P8M8S8VR8W8FK8DG8--KL8LED8DH8Y8M8IN8ED8Q8G8H8O8L8IT8V8V8PAD 5193  
Db 5406 PY8D8G8S8MIT8GI8Y8VE8K8DL8PE8GR8W8K8AS8FT8V8IT8-----Q8FT8V8S8GL8TED8Q 5451  
Qy 5194 MG8V8R8CLA8EN8MG8V8S8T8K8A8EL8R8V8DL8T8D8Y8D8TA8D8AT8ESS8Y8S8AQ8YL8SR8EQ8EG8TEST 5253  
Db 5452 R8Y8E8FR8V8IA8K8NA8GT8MS8-----K8PS8D8ST8-----G8P8ITA 5478  
Qy 5254 T8DE8QL8Q8PV8W8---E8EL8DL8Q8V8AP8GT8RL8AK8F8QL8K8V8K8Y8P8AP8RL8Y8W8FK8D8Q8P8T8ASA8H8I8MT8G 5311  
Db 5479 K8DE8V8EL8P8R8IS8MD8P8R8D8T8I8V8V8N8AG8ET8FL8E8AD8V8H8K8PL8T8I8E8W8L8RG8D8K8E8E8S8AR8C8I8KN 5338  
Qy 5312 KK8IL8HT8LE8I8IS8V8T8R8ED8S8Q8V8A8Y8IS8N8AM8GA8Y8S8AR8LL8VR8GP8DE8PE8K8P8AS8DV8HE8Q8LV8PP 5371  
Db 5539 T8DF8K8ALL8IV8K8DA8IR8ID8G8Q8Y8IL8R8AS8N8V8AG8S8K8S8FP8V8V8K8V8LD8R8P8GP8PE8GP8VQ8-----VT8G 5592  
Qy 5372 R8ML8ER8FT8PK8V8K8G8SS8T8F8SV8K8VB8--GR8PV8T8V8W8L8R8E8E8A8ER8G8VL8-W8IG8D8T8PG8Y8TV8ASS 5428  
Db 5593 VT8CE8K8CT8-----LT8W8P8PL8Q8D8G8SD8IP8--HY8V8VE8K8R8E8T8SL8AW8T8-----V8VASE 5634  
Qy 5429 AQ8Q8SL8V8LL8DV8GR8O8HQ8TY8T8CI8ASN8--AA8GO8AL8CS8AS8I8-----HV8S8GL8PK8-V8EE8Q8EK8V8K 5479  
Db 5635 V8T8NS8L8K8IT8K8LE8GE8Y8IF8R8IM8V8N8KY8G8V8GE8PL8S8AP8VL8M8KN8P8V8P8G8PK8S8LEV8T8N8IA8K 5694  
Qy 5480 B8AL8IS8T8F8T8Q8GT8Q8AIS8AG8LE8T8AS8F8AD8L8G8OR8KE8E8PL8A8E8AL8G8H8LS8LA8E8V8G8EE8F8O8KL 5539  
Db 5695 DS8MT8V8C8W8-----NR8P8DS8G8-----GS8-----I 5712  
Qy 5540 TS8Q8ITE8-----MV8SA8KIT8Q8AK8L8Q8VP8G8DS8D8D8SK8-TP8S8AP8RH8GR8SR8SS8IQ8E 5687  
Db 5713 T8GY8I8VE8K8R8DR8S8IR8W8IK8CN8K8R8V8T8DL8R8F8V8T8GL8T8H8E8Y8E8FR8V8SA8EN8A8AG8E8SP8A8---5769  
Qy 5588 SS8SE8SD8G8A8GE8IF8DI8Y8V8T8AD8YL8-----PL8G8AB8-----Q8D8AIT8L8REG8Q8-----5627  
Db 5770 -----TV8Y8K8AC8DP8V8FK8P8GP8PT8N8A8H8V8VD8T8TK8N8S8IT8L8AW8K8P8I8D8G8S8E8VL 5814  
Qy 5628 -Y8VE8LD8A8A8H8L8R8VL8V8T8K8P8T8K8S8PS8Q8W8S8P8AY8LD8R8RL8K8LS8---P8E8W8A8E8A8E8P8FG 5682  
Db 5815 G8Y8I8E8I8CK8ADE8E8Q8I8V8TP8QT8GL8KAN8R8FE8-ISK8L8IE8Q8E8K8IR8V8C8AL8NK8VL8G8E8A8A8V8PG 5873

5683 EAVSEDEYKARLSSVIOEL-LSSQAFVBELOFLQSHLQHLRCPHPIAVAGQKAVIF 5741  
5874 TVKEDKLEA-----PELDDSE-----LRKGVVRAAGSARIHIPKGRPTPDI 5918  
5742 RNVRDGRFHS-----SFLQ-ELQOCDTDDVAMCFINQQAFAEQYLEFLVGRV-QA 5791  
5919 TWSREGEFTDKVQKGVNFTQISINDCDR-NDAGKYIIVKLENSGKTAFVTKVLD 5777  
5792 ESVVSTAIQFYKKAEEALLAGDPQPPPLQHYLEQPV-----ERVQYQALLKELI 5847  
5978 PGPQNLAKEVKK---DSAVLWEP-----PIIDGAKVRYVIDKRE 6020  
5848 RNKARNRQNCALLQAVAVSALPQR-----AENKLHVSIMENYGP 5888  
6021 R-----KAVANVSKNKTTFKVENLTEGAIYFRVMAENFEGVGPVETVD 6067  
5889 TLEALGEPFROGHFVME-GAPGARMWPKNHRHVLFRNLHVLICKPRRDSRTDTSVYF 5947  
6068 AVKAAEPSPKGTLDVDSQTSASLWME-----KPEHGGSRVLGVV- 6110  
5948 RNMKLSIDLNDQVEGDDRAFEVWQEREDSV-----RKYLLQARTAIKSSW 5995  
6111 -----VEMQPKTEKWSVVAESKVCNAVVTGLSSGHEQYFRVKAINEGKS 6156  
5996 VKEICGIGIQRALPWRPDPDEELADCTAELGETVKLACRVCTGPKFVLSWKDGRAVQ 6055  
6157 DPRVLGPEV-IAKOLTIQSPKLPFKRYSVOAGDLKIEIPVIGRPRPEIFVWKGPELR 6215  
6056 VDPHILIEDPGSCALLDLSLTGVDGQVYCFEASAGNCS-TLGRILVQVPPFRVNVK 6114  
6216 -QTFVNVVEETATSTILHIKESKDDFGKYITATNSAGTATENLSVILEKPGPPVGPV 6274  
6115 R-----ASFPV-EEDAQFT---CTH-----EGAPYQIRWYKDAL-----LTTG 6151  
6275 RFDEISADVFVLSWEPAYTGCCQISNYIVEKRDITTTTHIVSATVARTIKVTKLKTG 6334  
6152 NKFO-TLSBPRSGLL-----VLV-----IRAAKEDLGLYECELNVRL 6188  
6335 SEYQPRIVAEIRYKSTSLDSKPIVYQPFKEPGPGTFTVTSVRQMLVQWHEPNDG 6394  
6189 GSARASAEIRIQSPMLQAOECHRRLV-----AAVEDTTL-----ERADQEVTSVLK 6236  
6395 GSKVLGYHL-----EQEKNSILWVKNVNTLIQDTRFKTGLDEGLEVEFFKVSAAE 6444  
6237 RLLGPKAPGPTGDLTGPGCPGAPALQETGSPVVTGTSEAPVPRVPQPLLEGPE 6296  
6445 NVIG-----IASLAKCPNAFVA-----RDFCDPP----- 6468  
6297 QPEAIARAQEWTPVIRMEGRAMPGAG--TGELL-----WDVHSHV-VRET----- 6339  
6469 GRPEALVITRN-NVTLKWKPKPAYDGSKITGVIVEKDLPGRWKMSFTNVLETFVTS 6527  
6340 ---TQRTYTYQADITHA-----RPPSMQVTIED-----VQAQTGGTA 6374  
6528 GLVPEQRYEFRVIARNAAGNLSEPSSESGAITARDEIDAPNASLDPKYKDVIVVHAGETP 6587  
6375 QFEALIEGDPQPSVWYKDSVOLVST-RLSQOQEGTYSVLVIRVASKADGAVTCLAQN 6433  
6588 VLEADIRGKPIEDVWVWLDKGKELETTARMEIKSTIQKTLVVKDCIETDGGQVYVLSLN 6647  
6434 TGGQVLCRALLVLGDNEPD 6454  
6648 VGGTKSLPITVAVLDRPGPPE 6668

RESULT 13  
S68235  
myosin-light-chain kinase (EC 2.7.1.117), 210K, nonmuscle - chicken  
N:Contains: myosin-light-chain kinase, 108K, smooth muscle; telokin  
C:Species: Gallus gallus (chicken)  
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999  
C:Accession: S68235; A37099; E44389; A44389; S28227; S78216; A35093; S11652  
R:Watterson, D.W.; Collinge, M.; Lukas, T.J.; Van Eldik, L.J.; Birukov, K.G.; Stepanova

FEB5 Lett. 373, 217-220, 1995  
A>Title: Multiple gene products are produced from a novel protein kinase transcription  
A:Reference number: S68235; MUID:96033976; PMID:7589469  
A:Accession: S68235  
A>Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-1906 <WAT>  
A:Cross-references: EMBL:X52876; NID:G992992; PIDN:CAA37056.1; PID:G992993  
R:Shoemaker, M.O.; Lau, W.; Shattuck, R.L.; Kwiatkowski, A.P.; Matrisian, P.E.; Guerra-  
J. Cell Biol. 111, 1107-1125, 1990  
A>Title: Use of DNA sequence and mutant analyses and antisense oligodeoxynucleotides to  
activity  
A:Reference number: A37099; MUID:90361738; PMID:2202734  
A:Accession: A37099  
A:Molecule type: mRNA  
A:Residues: 649-1906 <SHO>  
A:Cross-references: EMBL:X52876  
R:Collinge, M.; Matrisian, P.E.; Zimmer, W.E.; Shattuck, R.L.; Lukas, T.J.; Van Eldik,  
Mol. Cell Biol. 12, 2359-2371, 1992  
A>Title: Structure and expression of a calcium-binding protein gene contained within a  
A:Reference number: A44389; MUID:92236611; PMID:1373815  
A:Accession: B44389  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1695-1906 <COL>  
A:Cross-references: GB:M89284; NID:G212237; PIDN:AAB53767.1; PID:G212238  
A:Accession: A44389  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1750-1906 <CO2>  
A:Cross-references: GB:M89283; NID:G211371; PIDN:AAA48647.1; PID:G211372  
R:Yoshikai, S.I.; Ikebe, M.  
Arch. Biochem. Biophys. 299, 242-247, 1992  
A>Title: Molecular cloning of the chicken gizzard telokin gene and cDNA.  
A:Reference number: S28227; MUID:93073972; PMID:1444462  
A:Accession: S28227  
A:Molecule type: mRNA  
A:Residues: 1750-1906 <YOS>  
A:Cross-references: EMBL:X96655; NID:G212744; PIDN:AAA49083.1; PID:G212745  
A:Accession: S78216  
A:Molecule type: DNA  
A:Residues: 1750-1906 <YOW>  
A:Cross-references: EMBL:M96987  
R:Olson, N.J.; Pearson, R.B.; Needleman, D.S.; Hurwitz, M.Y.; Kemp, B.E.; Means, A.R.  
Proc. Natl. Acad. Sci. U.S.A. 87, 2284-2288, 1990  
A>Title: Regulatory and structural motifs of chicken gizzard myosin light chain kinase  
A:Reference number: A35093; MUID:90192792; PMID:2315320  
A:Accession: A35093  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 935-1438, 'Q', 1440-1906 <OLS>  
A:Cross-references: GB:M31048; NID:G212660; PIDN:AAA49069.1; PID:G212661  
R:Guerrero Jr., V.; Russo, M.A.; Olson, N.J.; Putkey, J.A.; Means, A.R.  
Biochemistry 25, 8372-8381, 1986  
A>Title: Domain organization of chicken gizzard myosin light chain kinase deduced from  
A:Reference number: A25810; MUID:87157587; PMID:3030394  
A:Accession: A25810  
A:Molecule type: mRNA  
A:Residues: 1258-1438, 'Q', 1440-1906 <GUE>  
C:Genetics:  
A:Introns: 1735/3; 1779/1; 1819/1  
C:Superfamily: myosin-light-chain kinase, nonmuscle; fibronectin type III repeat homolo  
C:Keywords: alternative initiators; ATP; calmodulin binding; phosphoprotein; phospho  
F:542-599/Domain: immunoglobulin homology <IMM1>  
F:935-1906/Product: myosin-light-chain kinase, 108K, smooth muscle (from 5.5kb transcri  
F:1098-1158/Domain: immunoglobulin homology <IMM2>  
F:1451-1708/Domain: protein kinase homology <KIN>  
F:1459-1467/Region: protein kinase ATP-binding motif  
F:1750-1906/Product: telokin (kinase-related protein KRP) (from 2.7 kb transcript) #sta  
F:1808-1869/Domain: immunoglobulin homology <IMM3>

Query Match 2.5%; Score 1012; DB 1; Length 1906;  
Best Local Similarity 19.7%; Pred. No. 1.1e-22;



QY 6511 R-ERDILAAALSHPLVTGLDQFETKRLTILILELCSSEELDLRYKGV-VTEAEVKVI 6568  
D 1496 RDEISWNCUHLHPKLVOCDAFEKANIVMWLEWVSGELFERILDEDFELTTEREIKYM 1555  
QY 6569 QQLVEGHLHSHGVLHLDIKPSNLMVHPAREDIKICDFGPAQNTIPAELOPSQYSGPE 6628  
D 1556 RQISEGVEYIHKQIVHLDLKPENIMCVNKTGTSIKLIDFGLARRLESAGSLVLEGTPE 1615  
QY 6629 FVSPETIQOQPVSEASDIWAGVTSILSLTSCSPFAGESDRATLLNVLEGRVSWSPMAA 6668  
D 1616 FVAPEVINYBPIGYETDMSIGVICYILVSGLSFPGDNDNETLANVTSATWDFDDEAFD 1675  
QY 6689 HLSDDAXDFIKATIQRAPOARPSAAQCLSHFWFLKSPABEAHFINTKOLKFLARSRWQ 6748  
D 1676 EISDDAXDFISNLKDKMKSRLNCTQCLQHPWLQKDTKNMEAKLSKDRMKYWARWKQ 1735  
QY 6749 RSLMSYKSLVMRSIPELLRPPDPSPLGVARHLCRDTGGSSSSSSSSSSSSSSSSSSSS 6808  
D 1736 KTGHAVRAI-----GRSSMAMISGMSGRKASGSS----- 1765  
QY 6809 LPPSPVTHSPHLLPRGFLRPSA 6830  
D 1766 -PTSPINADKVENEDAFLEVA 1786  
  
RESULT 14  
T34416  
hypothetical protein F12P3.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34416  
R:Fulton, B.; Weidmann, P.  
submitted to the EMBL Data Library, July 1998  
A:Description: The sequence of C. elegans cosmid F12P3.  
A:Reference number: Z21521  
A:Accession: T34416  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2783 <FUL>  
A:Cross-references: EMBL:U80022; PIDN:AA25886.1; GSPDB:GN00023; CESP:F12P3.2  
A:Experimental source: strain Bristol N2; clone F12P3  
A:Genetics:  
A:Gene: CESP:F12P3.2  
A:Map position: 5  
A:Introns: 45/3; 90/3; 451/3; 509/1; 2313/3; 2341/3; 2378/3; 2414/2; 2453/3; 2474/2; 252

Query Match 2.3%; Score 963.5; DB 2; Length 2783;  
Best Local Similarity 20.3%; Pred. No. 4.4e-21;  
Matches 667; Conservative 448; Mismatches 1225; Indels 945; Gaps 145;

QY 67 LTILDLALGDSQYVCRANAIGFAAAGLVQV-DAEACAEQAPHLRLRSTIRVREGS 125  
D 14 LTINSIDSDQGEFALKIKRGCEDKYAIGIQVTDPRPAAPGKPAVED-QNVDSVRLRWAA 72  
QY 126 EATFCRVGSP-----RPAVSWSKDGRRLGEPDGPVRVVEELGEASALRIRAAR 175  
D 73 PT-----NDGSPVRNYTEMCTEKTKTWK-----AEVTKQAFITFLNVL 113  
QY 176 PRDGTGYVRANPIGASAAALVDSDAADTASRPGTSTAALLAHLORREARBAEGA 235  
D 114 PGESYRFRVRADNTFG-----QSEPSD-----ESELVVVKVNSR----- 147  
QY 236 PASPSTGTRCTVTTEGKHARLSCVVTGPKETVWKDGLQVTEGR----- 282  
D 148 -----VVEPKKKEVVKVEQSDYDERVAKDSEPSYKTI 182  
QY 283 -RHVYVEDAQENFVLKILPKCSQDRGLYTCTASNLVGQTSYSLVWVEPAPVFKRLQD 341  
D 183 DIHRLPNDLQAKYIIIEELGKAGYGVYRAT-EKATGKTAAKMKVQVR-PGVKKENVIHE 240  
QY 342 LEVREKESATFLCEVQOPSTEAWKKESTRLWASAKYGLIEEGTERRLTVRNVSDDDA- 400  
D 241 ISMMNQLHHEKLLNLHEA-----FDWGNEMWLIBEFVSGGELFKILEDSDLSMEBEVR 294

QY 401 -----VYICETPEG-----SRTVAEL-AVQGNLLRL-PRKTAVRVGD 436  
D 295 DYMHOILLGVSHMHKQIVHLDLKPENILLKAKSNELKIIDFLGARKLDPKSKVLFG 354  
QY 437 TAMFCVEILAVPGVPHVHNRQEEVAVGGRVAISA-----EGTRHTLIISQCCLEDV 487  
D 355 TPEFCAPEVNYQPVGLSTDMWTGVI SYVLLSGLSPLGLSDSDDETLANVSASDWDFDP 414  
QY 488 G--QVAFMAGD--C-----QTSTRFCVSAAPKPPLOPPVDPV-VKARMESV--IL 531  
D 415 SWDDVSDLAADKIDICMLTKDKKRNVSQDALBPHMITKQPKLDKSGVPARQKRNFLSLK 474  
QY 532 SMSPP--PHGE-----RPVTIDGYLVEKKKLTGTYTWIRCHEAEWATBELTVDVAE 581  
D 475 RNSDOLLPTGLAKRGAIFRRLTMDGVF----- 502  
QY 582 EGNQFRVSALNSFGSPVLEFPGTVHLAPLAVRTPKAVQAVBG-----GEVTFVSVDLTV 638  
D 503 ERNIAPDTDA-----AP--SVKQLEDIVANVGDLIATLSCDVDGVP 542  
QY 639 ASAGWFLDGQALKASSV--YEIHCDRTRHTLTIREVPASLHGAQLKVFVANGIESI---R 694  
D 543 SPKVQWYKDDKELTVPSMKYDSFYNEGLAELTVKNI VESDAGKYTCRATNDLGSIMTHAK 602  
QY 695 MEVRA-----APGLTANKPPAAAAREVL-----ARLH-----BEAQL----- 726  
D 603 LSYKADKKKKKSKTSPAVIEKKDKORTSKVVVIEEMIDMPNPFHLLQDDAKIGEPKI 662  
QY 727 --LAELSDQAAAATVTKDGRITLS-PGPKYEQVQASAGRRVLLVRDVARDDAGLYECVSRGG 783  
D 663 LVVNTTLPETVDMVHNGEHSINDSNVLRKDKRGYELHLSVDSVTDGKWKAVG--- 719  
QY 784 RIAYQLSVQGLARFLHKOMAGSCVD-----AVAGGPAQFECETSEAHV----- 826  
D 720 -----KNAFGECESEAKLTWVDPQYAPSPFGKQSDVKCESDILKLE 763  
QY 827 -----HVHWYKDMELHSGRFLQEDVGT-RHRLVAATVTQDEGTYSK-----RV 872  
D 764 VNTQANPAPEINFRNESIEHSQHRLQFDGSGNYSLTIIDAIAEDSGEYKCVAKNKI 923  
QY 873 GEDSVDRLRVSE-----PKVYFAKEQLARKL-----QAEAGASATLSCEVA-QAOTEV 921  
D 824 GKAAHTVCCVRIBELLSKSKKIDGSKAPRFRMQLPTPREVPQCADITLVCSVSGTPHPNI 883  
QY 922 TWYKDGKLLSSSKVCWEATG-CTRRLVVOQAGQADAGEYSCFA---GGORLSFHLDVKE 977  
D 884 KWTKDDKPIDMSNKQVRHENGVCVTLHII--GARDDDQGRYVCEAENIHGVAOSF--SVVE 939  
QY 978 PKVYFAKDVYAHSEVQ-----AEAGANATLSCEVAQAQAEVMMYKDGKKL--SSSLKV 1028  
D 940 IKEAVDKDVRPKFLEPLVNCSTCEGNMVECCVTGKPIPTITWYKDLGLKIENRMLQ 999  
QY 1029 HVEAKGRRRLVVOQAGKTDAGDYSCAARGQVRSFELHT-----EPKMF 1075  
D 1000 YTRDKGV-SRLNIMNVMNDGDEYTCFAVNSLKGDPHTCTVKVDMGLSKRLTTPVRSKS 1058  
QY 1076 KEOS-----VHNEVQ-----AEA-----GASAMLSCEVAQAOT-EVYTWKDGKLLS 1116  
D 1059 RSRSRSPSVGGIQRPPVVTREPLADATVTEGNRELLEVEVDGFPPTIEWYHDGKLVAE 1118  
QY 1117 SSXVGMVEKCTGTRRLVLPQAGKADAGEYSCAARGQVRSFHLHITTEPKGVFAKE----- 1169  
D 1119 SRTLRTYFDGRVAFLLKIYEAEHEHNGQYVC-----KVSNKLGAETRAIVVVEAPDAEH 1173  
QY 1170 -----QSVHNEVQAEAGTTAMLCE-VAOPOTEVTWYKDGKKL--SSSSKVRMEVK 1217  
D 1174 VTQMPFTVKLQDVVLKTAGETATFTCCQSYANPAQVWHLNGKALQOIKSNYKRL-FD 1232  
QY 1218 GCTRLVVOQVQKADAGEYSCAAGG-----RVSFQLHI--TEPKAVFAKEQLVHNEVRTE 1271  
D 1233 DNTATLVENVDELCTYTAVANNGFQVHTSAQLTISGEAKKIAASLPYFIILKPK 1292





**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 11:05:23 ; Search time 82 Seconds  
(without alignments)  
5059.695 Million cell updates/sec

Title: US-10-077-130-5

Perfect score: 41273  
Sequence: 1 MDQPFSGAPFLTRPKAFV.....RNREKRALLYKHNLAQVR 7968

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2486	6.0	6632	1 UN89_CABEL	O01761 caenorhabdi
2	1074	2.6	4391	1 PGBM_HUMAN	P98160 homo sapien
3	1023	2.5	1914	1 KMLS_HUMAN	Q15746 homo sapien
4	1012	2.5	1906	1 KMLS_CHICK	P11799 gallus gall
5	909	2.2	3038	1 TRIO_HUMAN	O75982 homo sapien
6	767	1.9	3707	1 PGBM_MOUSE	Q05793 mus musculu
7	645.5	1.6	1176	1 KMLS_BOVIN	Q28824 bos taurus
8	622	1.5	1147	1 KMLS_RABIT	P29294 oryctolagus
9	586	1.4	1270	1 MYPC_MOUSE	O70468 mus musculu
10	572.5	1.4	1709	1 SN_HUMAN	Q9b222 homo sapien
11	571	1.4	1274	1 MYPC_HUMAN	Q14896 homo sapien
12	532	1.3	1694	1 SN_MOUSE	Q82230 mus musculu
13	528.5	1.3	1131	1 MYPF_CHICK	P16419 gallus gall
14	520	1.3	1142	1 MYPF_HUMAN	Q14324 homo sapien
15	519	1.3	3375	1 UN52_CABEL	Q06561 caenorhabdi
16	506.5	1.2	1271	1 MYPC_CHICK	Q90688 gallus gall
17	499.5	1.2	607	1 KML2_RABIT	P07313 oryctolagus
18	495.5	1.2	595	1 KML2_HUMAN	O9h1r3 homo sapien
19	489	1.2	448	1 DAK3_MOUSE	O94784 mus musculu
20	489	1.2	448	1 DAK3_RAT	O88764 rattus norv
21	480.5	1.2	609	1 KML2_RAT	P20689 rattus norv
22	475.5	1.2	370	1 DAK2_MOUSE	Q9vdf3 mus musculu
23	471	1.1	1432	1 DAK1_HUMAN	P53355 homo sapien
24	469	1.1	1442	1 DAK1_MOUSE	O80ye7 mus musculu
25	467	1.1	1141	1 MYPS_HUMAN	Q00872 homo sapien
26	466.5	1.1	454	1 DAK3_HUMAN	Q43293 homo sapien
27	465.5	1.1	370	1 DAK2_HUMAN	Q9uik4 homo sapien
28	464.5	1.1	1050	1 ULK1_HUMAN	OY5385 homo sapien
29	462.5	1.1	2012	1 DSCA_HUMAN	O60469 homo sapien
30	460.5	1.1	438	1 KMLS_SHEEP	O02827 ovis aries
31	456.5	1.1	1465	1 MYM2_HUMAN	P54296 homo sapien
32	451.5	1.1	1051	1 ULK1_MOUSE	O70405 mus musculu
33	451.5	1.1	1451	1 MYM1_HUMAN	P52179 homo sapien

#### ALIGNMENTS

##### RESULT 1

```

UN89_CABEL
ID UN89_CABEL STANDARD; PRT; 6632 AA.
AC
001761; Q17362;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR C09D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benlian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT assembly, encodes a giant modular protein composed of Ig and signal
RT transduction domains";
RL J. Cell Biol. 132:835-848(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Le T.T., Wilson R.;
RT Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Structural component of the muscle M-line. Myofilament
CC lattice assembly begins with positional cues laid down in the
CC basement membrane and muscle cell membrane. UNC-89 responds to
CC these signals, localizes, and then participates in assembling an
CC M-line.
CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 5 RCSD domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U33058; AAB00542.1; -
CC EMBL; AF003131; AAB54132.2; -
CC PDB; 1FHO; 20-DEC-00.
CC WormPep; C09D1.1; CE30426.

```

```

34 451 1.1 371 1 S17B_RAT Q91xs8 rattus norv
35 447.5 1.1 397 1 S17A_RABIT Q9gm70 oryctolagus
36 445.5 1.1 414 1 S17A_HUMAN Q9uee5 homo sapien
37 441.5 1.1 374 1 KCCL_EAT Q63450 rattus norv
38 440 1.1 424 1 KPSH_HUMAN P11801 homo sapien
39 439.5 1.1 1450 1 MFSF_CHICK Q02173 gallus gall
40 438.5 1.1 372 1 S17B_MOUSE Q8Dg48 mus musculu
41 437.5 1.1 374 1 KCCL_MOUSE Q81ys8 mus musculu
42 432 1.0 794 1 K111_HUMAN Q8tdc3 homo sapien
43 427 1.0 372 1 S17B_HUMAN O94768 homo sapien
44 426 1.0 370 1 KCCL_HUMAN Q14012 homo sapien
45 426 1.0 1666 1 MYM1_MOUSE Q62234 mus musculu

```





Db	2649	-----PLKVAPIVPEFETSLCDATREGDTLKLRAVLLGEPFVV---S	2690	Db	3685	KPLFTPEPLKETFANVEGDVTVLECKVKNESHPIQIKFFKNDQPVBIGQHMOLVLEDGMI-K	3743
Qy	2412	WLLGKTVLQAGNVGLQEGVHRLMLRRTCSWTGVPVHFTV---GKSRSSARLVVSDI	2467	Qy	3329	LOIRGLAMVDAEAYSC-----VCGERTSASLTI	3357
Db	2691	WYNGKKBESQNIKIHSKGTIVTITKIDTCDYSGQVCEAINEYKATSEATL-----	2745	Db	3744	LTIQNAKKEDVGAYRCEAVNVAGRANTNADLKIQAACKVEHVHDESQLEIGQFETVG	3803
Qy	2468	PWLTRPLEP-----KTG-RELQSVVSCDFRPAPKAVQVQKDTPLSPSEKFK	2515	Qy	3358	-----RPMFAHFTGLRHQESIEGATATLRCELSKAAP--VEMRK-GRESLRDG	3403
Db	2746	LVLPRGEPDFLEWLSNVRARTGTVHKVVF"GDPKP---SLTWYINNKEILNSDLYT	2801	Db	3804	DTASSKDTGTGAP--EFVELLRSCVTVEKQQAIIKCKV-KGEPRPKIKWTKEGVEVMSA	3861
Qy	2516	MSLQGMALPILRLMP-----ADAGVYRQCA-----GSAHSTSTVTEARE	2557	Qy	3404	DRHSRLQRQGAVCCELQICGLAVADAGEYSCVCEERTSA-----TLTVKALP-----AK	3451
Db	2802	IVTDDKTSTLTFNSFNDVHVGEIICAENDAGEVSTANMITVTSDFMESSESAQAE	2861	Db	3862	RVAEHKDDGTLTLTFDNTQADAGEYRCEAENEYGSATWTEGPIIVILEGAPKIDGAPD	3921
Qy	2558	-----VTVTGP-----LQDAEATEEGWASFCELSHEDEE---VE	2589	Qy	3452	FTEGLRNEEAVEGATAMWCELS-KVAP-VEMRKGPNLRDGDRIILR--QEGTRCELOI	3507
Db	2862	FVGDDLTEDESLEBEMHRTPTPVNAKPFITKIDTKAKKGHSAVFECVV--PDTKGVCCK	2919	Db	3922	FLQPVKPAVVTGETAIVLEGKISGPKPSVKWYKNGBELKPSDRVKIENLDDGTQ-RLTV	3980
Qy	2590	WSLNGMPLYNDSFHEIHSKGRHT-----LVLKSIQADAG-----IVRASSLV	2634	Qy	3508	CGLAMADAGEYLCVCGQE---RTSATLITRALPAR---FIEDVKNQEAAREGATAVLQC	3559
Db	2920	WLKDGKEI-----ELIARIVQRTGPEGHI TOELVLDNVTPEADAGKYTCIVENTAGD	2973	Db	3981	THAKLDDMDVRCESASNEFGDVMSDVLTVKE-PAQVAPGFFKELSAIQVKEETAKFEC	4039
Qy	2635	STASRLV-----RVKPVVFLKALDDLSAEBRGTLALQCEV-SDPEAHVVRKDGVO	2685	Qy	3560	ELNSAAP-VEMRKGSETLRDGR--YSLRQDGTCELOIRGLAMADTGEY-----SCVCGQ	3612
Db	2974	TCEATLVIESLEKKSEKKAPEFIVALQDTKTKSEKVLCKVIGEPKPSVWLHDKNT	3033	Db	4040	KVSGTKPDVKWFKDGTPLKEDKRVHFESTDDGTQ-RLVIEDSKTDDOGNRIEVSNDAGV	4098
Qy	2686	LGPSDKYDPLTAGTRGLVHVDVSPEDAGLYTCHV-----GSETRARVRVHDLHVGITR	2741	Qy	3613	ERTSAMLTVRALP---IKFTGLRNEEATEGATAVLRCCL-SKMAPVEMWKGHTLARDG	3668
Db	3034	I-TOESITVESVEGVERVITTSLSHQGKYCIAENTEGTSKTEAPLTVQGEAPVTFKE	3092	Db	4099	ANSKVPLTV--VPSETLKIKKGLTDVNTQTGKILLSVEVEGKPKTKVWYKGTBITVSSQ	4156
Qy	2742	LKTMVLGEGSCFECVLSHESADPAWTVGGKT-----VGSSRF-----QATRGKX	2791	Qy	3669	RHSLRQ-DGARCCELOIRGLVAEDAGEYLCMCKE---RTSAMLTVR-----AMPKFI	3717
Db	3093	LONKELSIGELVJSCVKGSPQPHVDYFSPSETTKVETKITSSRIAIEDQTNTHWM	3152	Db	4157	TTKIVQVTESEYKLEIESAEMSDTGAYRVVLSTDFSSESSATVTVTKAAEKISLPS-FK	4215
Qy	2792	YI-LVIREAAPS DAGEVVFVSRGJTSKASLIVRERPAAIKPLEBOWVAPGEDVELRCEL	2850	Qy	3718	EGLRNEEATEGDTATLWCEL-SKAAPVEMRKGHTLRDGRHSLRQDGSRCCELOIRGLAV	3776
Db	3153	VISQITKEDIVSKAIATNSIGTATSTSKITTKVEAPVPEQGLKTKTSKEKEBIKMEKV	3212	Db	4216	KGLADQSPVKPTPLVLEIEGPKDKVWYKNGDEIKGKVEDL--GNGKTRLTIPDQOE	4273
Qy	2851	SRAGTPHMLKDKRAIKRSQKDVV--CEGTWMLVIRGLASLKDAGCYTCEVEASKSTAS	2908	Qy	3777	VDAGYSVCV---CGQERTSATLTVRALPARFIEDVKNQEAAREGATAVLQCEL-SKAAPV	3831
Db	3213	GGSPADPEWFKDDRPVSDGNHEWKNKMPETGVFTLVKQAAATDAGKITAKANPAGTAE	3272	Db	4274	KDVGYSYVTAANEAGEIESKAKVNVSAKP-BIVSGLVPTTVKQGETATFNVKVGPKVGV	4332
Qy	2909	LHVBEKAN-----CFTEEL--TNLQVEEKGTAVFTCKTEHPAATVVRKGLLELRASGX	2960	Qy	3832	EVRKGSSETL-----RGDRYSRLRQDGTCELOIRGLHLSVADTGEYSVCV---CGQERTS	3880
Db	3273	SSAAEVTQSLEKFTFVRELVTTEVKINETATLSVTKGV-PDPSVEWLKDGQGPVQTDSS	3331	Db	4333	KWYKNGKIPDAKTNDGSGYS-----LEIPNAQVEDAADYKVVVNSNDAGDADSS	4383
Qy	2961	HQPSQ-EGL-TLRLTISALEKADSDTYTC-----DIGQAQSAQLLVQGRV--HIIDLE	3012	Qy	3881	ATLTVRAP-----OPVREPLQSLQAEBSGSTATLQCELSPTATVWVSKGLQLOANG	3933
Db	3332	HVIKVEGSGYSTIKDARLEDSGKYACRATNPAGEAKTEANPAVVKNLVPPPEFVEKLS	3391	Db	4384	AALTVKLADDGKDKVKPIVSGLIPPTTVKQGETATFNVKVGPKVKNKNGKEI-PNA	4442
Qy	3013	DVDVOEGSSATFCRISPANPEVHFLDKTPLHANEINEIDAQPG-GYHVLTLRLQALK	3071	Qy	3934	RRPELQGTAEVLQDLQREDTGEYTCGSOA-----TSATLTVTAAPVRLREHQOE	3989
Db	3392	PLEVKEKESTLSVKVGTGTPSPVEMFKDDTPISIDNVHVIQKTA VGSFSLINDARQ	3451	Db	4443	KAKDNGDG-SYSLTEPNAQLDDTADYKVVVNSNDAGDADSSAALTVKPLPGIAIVKGLDAE	4501
Qy	3072	DSGTIYF-----ERAGDQASALARVTEK--PSVFSRELTATTEGEDLTLVETSTCDI	3124	Qy	3990	VDEGCTAHLCCLSLSPAGASVEMRKGSLQLFCAKYQVMVQDG-AAAEALLVRGVEOEDAGDY	4048
Db	3452	DVG-IYSCRAEAGEALTANFNGIIRDSIPPEFTOKLREVBREQETLTKVTVIGTPV	3510	Db	4502	VPKGGKAVLQVETNKKPKIKWYKNGKEITPDSKRAQPSGSDNKPQLVIPAGDDDAEY	4561
Qy	3125	P-MCMTKDKGTLRGSARCOLSH-----EGHRAQLLITGATLQDSGRYKC-----EAGGA	3172	Qy	4049	----TCDTGT--QSMASLSVRVP--RPXFKTRLOSLEQETGDI-ARLCCQSDAESGAVV	4100
Db	3511	PNVEMFKDDKPI-----NIDNSHIFAKDEGSGHHT-LTIKQARGEDVGYTCKATNEAGEA	3565	Db	4562	KVVLTDGENTADSSCALTVKLPKPEKPI---IKGLEQVQVSIIGSPIKLETSGPSKTV	4618
Qy	3173	CSSSIVRVHAR---PVRQEAELKOLEVEGGAATLRCVLSSVAP-VKWCYGNVLRPGD	3228	Qy	4101	QWLKGEVHLAGPKYKEMSSQCATRELLIHOLEA---KDTGEYACVTGSGQKTA--SLRVT	4155
Db	3566	KTTANMAVQEBIEAPL-FVQGLKPYEVEQGPFAELVVRVEGKPEPEVKFKDGPVPIADN	3624	Db	4619	KWYKNGKELPGAAATIKIKIDDKNKYIEIPSSVVEDTGYKVEVANEAGSANSKGIT	4678
Qy	3229	KYSLRQEG--AMLEIVLRNLRPQDSGRYSC-----SFGDQTTSATL-----TVTAL	3272	Qy	4156	-BPEVTVIRGLVDAEVTADEVFECSVRAGATGVQVCLQGLPLQSNNEVTEAVRDGRI	4214
Db	3625	QHVIEKKGENSGHTLVIKDTNNAFGKYTCQATNKAQKDETVGELKIPKYSFEKQTAEEV	3684	Db	4679	VEPKITFLKPLKDKQSITEGENAEFVSVENTKPI-KWYKNGQBEIKPNSREIFIQKTDTK	4737
Qy	3273	PAQFIGKLRNKEATEGATATLRCELSKTA--PVEMRKGSETLRDGDRIYCLR--ODGANCE	3328	Qy	4215	HTLRLKGVTPEDAGTVSFHLGNHA-----SSAQLTVRAPEV---TILEPLQDVQLESGODA	4267
Db				Db	4738	YQLVTKNAVRRDADTYKTVLENTAGEAESQLTVKKAKAGLCKIVKGLEDOQVAVKAGKM	4797

QY 4268 SPQCLRSASGOEARWALGVLPLQANEMNDITVEQTLHLTLHLKVTLEDAGTVSFHV-- 4325  
D 4798 VFEVKI--QGBEPDVRMLRDNANVIGAGANAIEKIDDDTYRLIIISADLKAGAEVTVVIN 4856  
QY 4326 --GTCSSBAQKVTAKTNTVRGLNVEALEGEALFECQLSQPEVAHAHTWLLDDEPVRTS 4383  
D 4857 ESGRAKSDAKGEVDEKPEIPEVGLNIEIPEGGDDVFKVEVSAP--VRQVKWYKNDQEIKN 4915  
QY 4384 ENAEV-----VFENGRLHLLKRLRQDSCRVTFLAGDMVTSAFLT--VR 4428  
D 4916 SHLEAKKIPKYYELAINRAQLDDGADYKVVLSN-----AAGDCSSAALTIVK 4964  
QY 4429 GWRLEIPLKVAARAGAAQARFCTTISEAVPVGEAGWYINGAAVQPDSDMTVTADGSH 4488  
D 4965 PNVLKIVDGLKDVDVEEPQVELKVKV--EGIP--KVIKWKNGCQELKPD-----ADG-- 5013  
QY 4489 QALLRSQAQPHHAGEVTFACRDASARLTVLGLPDPDEDAEVVAHSSHTVTLSSWAAPMS 4548  
D 5014 -----FKPEEPESGE-----FSLTIP--SSKKS 5035  
QY 4549 DGGGLCGYRVEVEKAGATGWRQLCHELVPGCEVVDGLAPGETYRFRVAAGVPVGAPEV 4608  
D 5036 DGG--AYR-----VVLGNDKGEVY-----SGSVV 5057  
QY 4609 HLPQTVRLAEPKVPPOP--SAPESQVAAAGEDVSLELVVAA--GEVWHKGMERIQP 4665  
D 5058 H-----VKSAKSEPTSGANFLSPKLTDEVEEGMLTLQCTIAGPFPPEVWEKDGVLQK 5113  
QY 4666 GGRPEV--VSQGRQOMLVTKGTADQGEYHCGLAQGSICPAAATFOVALSPASVDEAPQP 4724  
D 5114 DDIRTMVALDGTATLRSKAKSDIGQYRV--TAKNEAGSATSCKTVT-----BQGEQP 5168  
QY 4725 SLP-----PEAAQEGDLHLLEALARKRMSREPTL-----DSISELPEED 4765  
D 5169 SKPKFVILPKTGAALPGDK-----KEFNKVRGLPKPTLQWFLNGIPIKFDRIITLDMAD 5224  
QY 4766 GRSQRLPOEAEEVAPDLSEGYSTADELARTGDADLSHTSDDESEAGTSPSLVTLKAGR 4825  
D 5225 GNYCLTIRDVRE-----EDFTLKCTAK-----NENGIDE-----TVCEPQQAGH 5265  
QY 4826 PGTSLASKVGAPAAPSVKPOQOEPLAAVRPPLGLDLSKDLGDPMSDKAAVKITQAAPKG 4885  
D 5266 -----DGSRDDLRYP----- 5276  
QY 4886 YKVRKEMKQOQPMFSTFGTEAQVGDALBLECVASAKADVRARWLKDGVELTDGRHHH 4945  
D 5277 -----PRFNVLWDRRIPVGDPMFIECHVDANPTAEVWFKDGKKEHTAHE 5324  
QY 4946 IDQLGDGTCSLIAGLDRADAGCYTCQVSNKFGQVTHSACVWVSGSESEASSGGELDD 5005  
D 5325 IRNTVDGACRIKIIPFEESDIGVMCVAVNELGQATCATVQVLELHVEEKR----- 5378  
QY 5006 AFRAARLHLRFTKSPAESVDSBELFLSADGPAEPPEPADWQTYREDEHFCIRFEAL 5065  
D 5379 --REYAPKNPLEDKTVNGQPIRLSKVD--AI PRASVW--YKDG-----LPLR 5424  
QY 5066 TEAQAVTRFOEMFATLIGVIEIKLVEGPRRVMCI--SK-----ETPAPVVRPEPLPSL 5119  
D 5425 ADSRTSIQYEDGCTATLAINDSTE--EDIGAYR--CVATNAGTINTSCSVNWKVQEV 5480  
QY 5120 LTSDAAPVFLTELQNEQVQDGYPSVFCVVTGQPMPSVRMPKDGKLEEDHMYNEDDQ 5179  
D 5481 KKEGEEPPFTKGLVDLWADRGSFTLKCAVTGDPFPFSIKWYRNGQLLRNGPRVTIETSPD 5540  
QY 5180 GGHQLIITAVPADMGVVECLAENSMGVSTKABLRVDLTSTDYDTAADATESSESYESAQ 5239  
D 5541 GSCSTVNSTNDEGIYRCAENAHGAKAKTOAHVQM-----ALGKTE----- 5585  
QY 5240 GYLSSREGEGETSTDEQQLPQVVEELRDLOVAPGTRLAKFQLKVKGYVAPRLYFMDQG 5299  
D 5586 -----KPRWDECKPKFILELSDMSVSLG--NVIDLECKVTGLPNPSVKWSKDG 5633

QY 5300 PLTASAHIRMTGK--KILHTLEIISVTREDSGQYAAVISNAMGAAYSSARLLVGRDEPE 5357  
D 5634 PLIEDSRPWSNEASGVVQLRIKNAIVHDEGYRCVATNENGA--TTKSFVR--MD 5687  
QY 5358 EXPASDVHQLVPPMLERFTPKVKKSSITFSVKVEGRPVPTVHHLREBA-----ERG 5412  
D 5688 DGLSGGVVTSQPPRFTLXMGDVRTTEGQPLKLECKVDASPLPEVMWYDGAIVTPSDRI 5747  
QY 5413 VLMIGDPTPGYTVASSAQOHSILVLLDVGRHQHGTCTIASNAAGALCSASLHVSGLPK- 5471  
D 5748 QISLSPD---GVA-----TLLIPSCVYDDGIRVVIATNPSGTAQDXGTATVKLPRD 5797  
QY 5472 -----VBEQEKVKEALIS-----TPL 5487  
D 5798 SGARRSADRVFDANKAPKMEPLENIRIPEXQSFRLCRCKSPGCDPKPTIKWFKDGERVFP 5857  
QY 5488 QGTTOAI--SAQG-----LETASFADLGGORKEEPLAAKEALGHLSLAEVTEFLQKLS 5541  
D 5858 YGRLOLIESPDGUCELVDSATRODAGYR--CVAENTYG----- 5895  
QY 5542 QITEMVSAKITQAKLVQPGD--SDEDSKTPSAPRHRGRSRPSSIOSSSESDGDAR 5598  
D 5896 -----SAR--TSCDNNVIRDRKPRDIDS-----SIREGKA--PGFTTPTLIRRAKPGDS- 5941  
QY 5599 GEIPDIYVVTADYLPGLAEQDAITLREGQYVEVLDAAHPLRLVRLTKPKSSPSRQGWYS 5658  
D 5942 -----VTFECLPFGNPPFSI-----KWL-----KDG--L 5963  
QY 5659 PAYLDRRLKLSPEWGAARPEFPGGAVSDEYKARLSSVIOELLSEQAQFVBELOFQSH 5718  
D 5964 ELFSDSEKIM-----EAAADG-----TQRLILSDVTFLESEGYT- 5996  
QY 5719 HLQHLERC-----PHVPIAVAGOKA--VIFRNVDRIDIRFHSFLOELQOCCDDDDVAMCFIK 5773  
D 5997 -----RCVATNEH-----GTASTKAEVIEGDRITIG----- 6022  
QY 5774 NQAAPQYLEFLVGRVQAESVVVSTAIQEFYKYAEALLAGDPSOPPPPLQHYLEQPV 5833  
D 6023 -----SRPLP----- 6027  
QY 5834 ERVQRYQALLKELIRNARKNRQNCALLEQAYAVVSALPORAENKLHVSIMENYPTGLEAL 5893  
D 6028 -----EYN 6030  
QY 5894 GBPIROGHFIVMEGAPGARPKHNRHVLPFRNLHVICKPRDRSDTDTVSVFRNMKL 5953  
D 6031 GBPEE-----CKPR-----IRGLYNN 6047  
QY 5954 SSIDLNDQVEGDDRAPEVMQEREDSVRYKYLQARTAIKSSWVKEICIGIQORLALPVMRP 6013  
D 6048 S-----IHE----- 6051  
QY 6014 PDPEELADCTAELGETVKLACRVGTGTPKPVISWYKDGKAV--QVDPHILIEDPDGSC 6070  
D 6052 -----GNVEMIVCATGIPPTVKWYKDGQEI VGDGPKRVITFDERGHH 6097  
QY 6071 ALILDSLTGVDGQVMCFASAGNCSTLGLKILV-----QVPPRFVNVKR 6115  
D 6098 HLVIYNASPDDEGEYSLEATNKLGSAKTEGSLNIIRPHIADADERRGMPFPFPGVRQLK 6157  
QY 6116 ASPFVEGEDAQCTCTIEGAPYQIRWYKDGALLTTGNKFPQTLSEPRSGLLVLVIRAAKE 6175  
D 6158 NKHFVNHMTPIFDCLVWGHPAPEVFNHNGKIVPGRIK--IQSCGGGSHALIILDTLE 6216  
QY 6176 DLGLYCEELVNRLGSGARASAEALRIQSPMLQAEQCHREQLVAAVEDTTLERADQVTSVL 6235  
D 6217 DAGEYATAKNHHGSSASSAVLDVTVPL-----DSIKFNGEIDVTPYL 6260  
QY 6236 KRLLG-----PKAPGSTGLDTPG-----PCPRGAP----- 6262  
D 6261 TEEYGPKLNTASLPTFPDRGPFKEVTHYLTLSWIPTKRAPPYQVTVYVIEIRELPE 6320  
QY 6263 ---ALQETQSQSPVT-----GTISE--APAVPP-----RVPQPLLHE 6293

```

Db 6321 KOWSLEETNIEPEYKVRNLELGSQYFRVRAENIYGISDPSASPSPRLMAPPQPVFDR 6380
QY 6294 -----GPEQSEPAIARAQE-----W-----TVPIRMEG----- 6316
Db 6381 RTNKVIPLELDYAEKALDMRYSEYACAPMSPGVKRYCAENDTLTIVINSGPDPD 6440
QY 6317 AANPAGAGTGLLWVHSHVRETTQRT-----YTYQAIDHTA----- 6354
Db 6441 IKWKFRG-----WDI-----DTSSPTSCKKVYTYGGSETTLAITGSKENVGQYQCFK 6489
QY 6355 -----RPPSMQVIEDVQACTGTAQ-----FEALIEDGPPSPVWTKDSV 6395
Db 6490 NDYDGAQONIMVLAIRNFQIPLWN-----KTFSSAQPMRMDVRVDGEPFELKWKKEWR 6545
QY 6396 QLVDSRLSQOQEGT-TYSLVLRHVASKDAGVYTCIAQNTGGQVLCKAELLVLGGDNEP 6454
Db 6546 PIVESSRIKFQDGPYLCSLIINDPMWEDSGIYSCVAVNDAGQATTCTVTVEAGDYND 6605
QY 6455 SEKQSHR-----RKLHSFYEVKEE 6473
Db 6606 VELPRRRTVIESRRVRELYEISEK 6629

RESULT 2
PGEM_HUMAN
ID PGEM_HUMAN STANDARD; PRT; 4391 AA.
AC P98160; Q16287; Q9H3V5;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Human basement membrane heparan sulfate proteoglycan core protein: a
DE 467-kD protein containing multiple domains resembling elements of the
DE low density lipoprotein receptor, laminin, neural cell adhesion
DE molecules, and epidermal growth factor."
DE protein precursor (HSPG) (perlecan) (PLC).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=92112994; PubMed=1730768;
RA Kallunki P., Tryggvason K.;
RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecules, and epidermal growth factor."
RL J. Cell Biol. 116:559-571(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Colon, and Skin;
RX MEDLINE=92235084; PubMed=1569102;
RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG2/perlecan): A chimeric molecule with multiple
RT domains homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor."
RL J. Biol. Chem. 267:8544-8557(1992).
RN [3]
RP SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJS1 TVR-1532.
RX MEDLINE=20553141; PubMed=1101850;
RA Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barral D.,
RA Beighton P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA Sanson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,
RA Hentati F., Fontaine B.;
RT "Perlecan, the major proteoglycan of basement membranes, is altered in
RT patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia).";
RL Nat. Genet. 26:480-483(2000).
RN [4]
RP SEQUENCE OF 1016-1470 FROM N.A.
RX TISSUE=Colon;
RX MEDLINE=91365376; PubMed=1679749;
RA Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA Yi H.F., Iozzo R.V.;

```

```

RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT short arm of human chromosome 1.";
RL Genomics 10:673-680(1991).
RN [5]
RP SEQUENCE OF 890-1396 FROM N.A.
RX TISSUE=Fibrosarcoma;
RX MEDLINE=92120660; PubMed=1685141;
RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA Tryggvason K.;
RT "Cloning of human heparan sulfate proteoglycan core protein,
RT assignment of the gene (HSPG2) to lp36.1--p35 and identification of
RT a BamHI restriction fragment length polymorphism.";
RL Genomics 11:389-396(1991).
RN [6]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE=94052171; PubMed=8234307;
RA Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT "Structural characterization of the complete human perlecan gene and
RT its promoter.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN [7]
RP CARBOHYDRATE-LINKAGE SITE ASN-2121.
RX MEDLINE=22660472; PubMed=12754519;
RA Zhang H., Li X.-J., Martin D.B., Aebersold R.;
RT "Identification and quantification of N-linked glycoproteins using
RT hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RL Nat. Biotechnol. 21:660-666(2003).
CC -!- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Found in the basement membranes.
CC -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -!- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel
CC syndrome (SJS1) [MIM:255800]; a rare autosomal recessive disorder
CC characterized by permanent myotonia (prolonged failure of muscle
CC relaxation) and skeletal dysplasia, resulting in reduced stature,
CC kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.
CC -!- SIMILARITY: Contains 4 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC -!- SIMILARITY: Contains 3 laminin IV domains.
CC -!- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 3 laminin G-like domains.
CC -!- SIMILARITY: Contains 4 EGF-like domains.
CC -!- SIMILARITY: Contains 1 SEA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62515; CA44373.1; -.
DR EMBL; M85289; AAA52700.1; -.
DR EMBL; AL445795; CAC18534.1; -.
DR EMBL; M64283; AAA52699.1; -.
DR EMBL; S76436; AAB21121.2; -.
DR EMBL; L22078; -. NOT_ANNOTATED_CDS.
DR PIR; A38096; A38096.
DR HSP; P00740; IEDM.
DR Siena-2DPAGE; P98160; -.
DR MIM; 142461; -.
DR MIM; 255800; -.

```







RT "Myosin light chain kinase in endothelium: molecular cloning and  
 RL regulation.";  
 RA Am. J. Respir. Cell Mol. Biol. 16:489-494 (1997).  
 RN [2]  
 RP REVISIONS.  
 RA Birukov K.G., Garcia J.G.N.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3A; 3B AND 4).  
 RX TISSUE=Umbilical vein;  
 RA Lazar V.L., Garcia J.G.N.;  
 RL MEDLINE=99216419; PubMed=10198165;  
 RT "A single human myosin light chain kinase gene (MLCK, MYLK).";  
 RL Genomics 57:256-267 (1999).  
 RN [4]  
 RP REVISIONS (ISOFORM 2).  
 RA Birukov K.G., Garcia J.G.N.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 923-1914 FROM N.A.  
 RX TISSUE=Hippocampus;  
 RA Potier M.-C., Chelot E., Pekarsky Y., Gardiner K., Rossier J.,  
 RL Turnell W.G.;  
 RT "The human myosin light chain kinase (MLCK) from hippocampus:  
 RL cloning, sequencing, expression, and localization to 3qcen-q21.";  
 RL Genomics 29:562-570 (1995).  
 RN [6]  
 RP SEQUENCE OF 1614-1914 FROM N.A.  
 RX TISSUE=Lung, and Placenta;  
 RA MEDLINE=20007838; PubMed=10536370;  
 RA Waterson D.M., Schavocky J.P., Guo L., Weiss C., Chlenski A.,  
 RA Shinsky V.P., Van Eldik L.J., Haiech J.;  
 RT "Analysis of the kinase-related protein gene found at human chromosome  
 RT 3q21 in a multi-gene cluster: organization, expression, alternative  
 RT splicing and polymorphic marker.";  
 RL J. Cell. Biochem. 75:481-491 (1999).  
 RN [7]  
 RP SEQUENCE OF 1456-1914 FROM N.A.  
 RX TISSUE=Placenta;  
 RA Waterson M.D.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: CALCIUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR  
 CC SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SERINE  
 CC IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT  
 CC FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS. CENTRAL  
 CC DETERMINANT IN THE DEVELOPMENT OF VASCULAR PERMEABILITY AND TISSUE  
 CC EDema FORMATION. IN THE NERVOUS SYSTEM IT HAS BEEN SHOWN TO  
 CC CONTROL THE GROWTH INITIATION OF ASTROCYTIC PROCESSES IN CULTURE  
 CC AND TO PARTICIPATE IN TRANSMITTER RELEASE AT SYNAPSES FORMED  
 CC BETWEEN CULTURED SYMPATHETIC GANGLION CELLS. CRITICAL PARTICIPANT  
 CC IN SIGNALING SEQUENCES THAT RESULT IN FIBROBLAST APOPTOSIS.  
 CC -!- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin  
 CC light-chain] phosphate.  
 CC -!- SUBUNIT: ISOFORM TELOKIN BINDS CALMODULIN.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=6;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=1; Synonyms=Non-muscle isozyme;  
 CC IsoId=Q15746-1; Sequence=Displayed;  
 CC Note=The smooth muscle isozyme and telokin are produced by  
 CC alternative initiation at Met-923 and Met-1761 of isoform 1;  
 CC Name=2;  
 CC IsoId=Q15746-2; Sequence=VSP\_004791;  
 CC Name=3A;  
 CC IsoId=Q15746-3; Sequence=VSP\_004792, VSP\_004794;  
 CC Name=3B;  
 CC IsoId=Q15746-4; Sequence=VSP\_004791, VSP\_004792, VSP\_004794;  
 CC Name=4;  
 CC IsoId=Q15746-5; Sequence=VSP\_004792, VSP\_004793;  
 CC Name=Del-1790;  
 CC IsoId=Q15746-6; Sequence=VSP\_004795;  
 CC Event=Alternative initiation;  
 CC

CC Comment=3 isoforms, 1/Non-muscle isozyme (shown here),  
 CC smooth-muscle isozyme and telokin, are produced by alternative  
 CC initiation at Met-1, Met-923 and Met-1761. Telokin has no  
 CC catalytic activity;  
 CC -!- TISSUE SPECIFICITY: SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES ARE  
 CC EXPRESSED IN A WIDE VARIETY OF ADULT AND FETAL TISSUES AND IN  
 CC CULTURED ENDOTHELIUM WITH QUALITATIVE EXPRESSION APPEARING TO BE  
 CC NEITHER TISSUE- NOR DEVELOPMENT-SPECIFIC. NON-MUSCLE ISOFORM 2 IS  
 CC THE DOMINANT SPLICED VARIANT EXPRESSED IN VARIOUS TISSUES. TELOKIN  
 CC HAS BEEN FOUND IN A WIDE VARIETY OF ADULT AND FETAL TISSUES.  
 CC -!- PTM: MLCK IS PROBABLY DOWN-REGULATED BY PHOSPHORYLATION.  
 CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.  
 CC -!- SIMILARITY: Contains 9 immunoglobulin-like C2-type domains.  
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U48959; AAC18423.2; -;  
 DR EMBL; AF069601; AAD15921.2; -;  
 DR EMBL; AF069602; AAD15922.1; -;  
 DR EMBL; AF069603; AAD15923.1; -;  
 DR EMBL; AF069604; AAD15924.1; -;  
 DR EMBL; X85337; CAA59685.1; -;  
 DR EMBL; AF096771; AAD51380.1; -;  
 DR EMBL; AF096766; AAD51380.1; JOINED.  
 DR EMBL; AF096767; AAD51380.1; JOINED.  
 DR EMBL; AF096768; AAD51380.1; JOINED.  
 DR EMBL; AF096769; AAD51380.1; JOINED.  
 DR EMBL; AF096770; AAD51380.1; JOINED.  
 DR EMBL; AF096774; AAD54018.1; -;  
 DR EMBL; AF096771; AAD51381.1; -;  
 DR EMBL; AF096769; AAD51381.1; JOINED.  
 DR EMBL; AF096770; AAD51381.1; JOINED.  
 DR EMBL; X90870; CAA62378.1; -;  
 DR HSSP; Q63450; 1A06.  
 DR Genew; HGNC:7590; MYLK.  
 DR MIM; 600922; -;  
 DR GO; GO:0004687; F-Myosin-light-chain kinase activity; TAS.  
 DR GO; GO:0004688; P-protein amino acid phosphorylation; TAS.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR000719; Prot kinase.  
 DR InterPro; IPR008271; Ser\_thr\_kin\_AS.  
 DR Pfam; PF00041; fn3; 1.  
 DR Pfam; PF00047; ig; 8.  
 DR Pfam; PF00059; pkinase; 1.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR PROSITE; PS00835; IG LIKE; 9.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_SF; 1.  
 DR Transferrase; Serine/threonine-protein kinase; Calmodulin-binding;  
 KW ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;  
 KW Alternative initiation; Alternative splicing.  
 FT CHAIN 1 1914  
 FT MYOSIN LIGHT CHAIN KINASE, ISOFORM NON-  
 FT MUSCLE ISOZYME.  
 FT MYOSIN LIGHT CHAIN KINASE, ISOFORM  
 FT SMOOTH-MUSCLE ISOZYME  
 FT MYOSIN LIGHT CHAIN KINASE, ISOFORM  
 FT TELOKIN.  
 FT  
 FT INIT MET 923 923  
 FT INIT\_MET 1761 1761  
 FT DOMAIN 33 122  
 FT DOMAIN 161 249  
 FT DOMAIN 414 503  
 FT DOMAIN 514 599  
 FT DOMAIN





DR EMBL; M31048; AAA49069.1; -  
DR EMBL; M14953; AAA69964.1; -  
DR EMBL; M96655; AAA49083.1; -  
DR EMBL; M88283; AAA48847.1; -  
DR EMBL; M88284; AAA53768.1; -  
DR PIR; S68235; S68235.  
DR PDB; 1CDL; 31-AUG-94.  
DR InterPro; IPR008957; FN III-like.  
DR InterPro; IPR003961; FN III.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003598; IG-C2.  
DR InterPro; IPR000719; Prot kinase.  
DR InterPro; IPR008271; Ser\_thr\_kin\_AS.  
DR InterPro; IPR002290; Ser\_thr\_kinase.  
DR Pfam; PF00041; fn3; 1.  
DR Pfam; PF00047; ig; 9.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SMO0060; FN3; 1.  
DR SMART; SMO0408; IGC2; 8.  
DR SMART; SMO0220; S TKC; 1.  
DR PROSITE; PS00835; IG LIKE; 9.  
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.  
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
DR transferase; Serine/threonine-protein kinase; Calmodulin-binding;  
KW ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;  
KW Alternative initiation; 3D-structure.  
FT CHAIN 1 1906  
FT CHAIN 210  
FT CHAIN 935 1906  
FT CHAIN 1750 1906  
FT CHAIN 1750 1906  
FT INIT MET 935  
FT INIT MET 1750  
FT DOMAIN 28 117  
FT DOMAIN 156 244  
FT DOMAIN 429 517  
FT DOMAIN 521 613  
FT DOMAIN 637 725  
FT DOMAIN 735 830  
FT DOMAIN 1084 1172  
FT DOMAIN 1225 1313  
FT DOMAIN 1330 1400  
FT DOMAIN 1453 1708  
FT DOMAIN 1794 1885  
FT NP BIND 1459 1467  
FT BINDING 1482 1482  
FT ACT SITE 1574 1574  
FT DOMAIN 1716 1728  
FT DOMAIN 1730 1749  
FT DOMAIN 1317 1364  
FT DOMAIN 1385 1402  
FT DOMAIN 660 1833  
FT REPEAT 660 676  
FT REPEAT 758 774  
FT REPEAT 1107 1123  
FT REPEAT 1817 1833  
FT DOMAIN 693 1866  
FT REPEAT 693 708  
FT REPEAT 791 807  
FT REPEAT 1140 1156  
FT REPEAT 1281 1297  
FT REPEAT 1851 1866  
FT DOMAIN 970 1226  
FT REPEAT 970 987  
FT REPEAT 999 1016  
FT REPEAT 1061 1078  
FT REPEAT 1209 1226

FT DOMAIN 1700 1763 CALMODULIN-BINDING.  
FT DOMAIN 1896 1906 POLY-GLU.  
FT MOD\_RES 1748 1748 PHOSPHORYLATION.  
FT MOD\_RES 1762 1762 PHOSPHORYLATION.  
FT CONFLICT 1439 1439 R -> Q (IN REF. 4).  
SQ SEQUENCE 1906 AA; 210445 MM; AD7D8A3B69E3363 CRC64;  
Query Match 2.5%; Score 1012; DB 1; Length 1906;  
Best Local Similarity 19.7%; Pred. No. 5.3e-27;  
Matches 488; Conservative 318; Mismatches 860; Indels 816; Gaps 76;  
4437 PLKNAVRAGAQAQRTLTSEAVPVGEASWINGAAVQDDSDTWTATDGGSHQ---ALLL 4493  
Db PPNIRVQLGATARFEGKV-RGYPPQITWYENGHPL-PEGDHVYV--DHSIRGIFSLVI 88  
QY 4494 RSAQPHHAGEVTFACRDVASARLTVLGLPPPEPAEVVAHSSHVTVTLSWAAPSDDGGG 4553  
Db 89 KGVQEGDSCKYTCEAANDGGVRQVTV-----ELTVEGNSLKYSLEPSAKTPGGR 138  
QY 4554 LCGYRVEVKEGATGQWLCHLVPGCEVWDGLAPGETYRFRVAAGVPVGAGEPVHLPTQ 4613  
Db 139 LSVPPVEHRPSIWGE--SPPKFATKPNV--VREGQTGRFSCKITG----- 181  
QY 4614 VRLAEPKPPVPPQPSAPESRQVAGEDVLSLEVVAAEAGEVIMHGMERIOPGRFEVVS 4673  
Db 182 -----RPQP-----QVTWTKGDIHLQONERFNFE 206  
QY 4674 QGRQQLVTKGFTABDQGEYHCGLAQGSICPAAATFQVALSPASVDEAPQP-SLPP--- 4728  
Db 207 KTGIGYLIQVIONVOLADAGIYCTVVSAGKASVABELTVQGPDKTDTHAQPLCPHPPTT 266  
QY 4729 ---EAAQEGDHLHLEALARKRMSREPTLDSISELPEEDGRSQRLPQEAEEVAPD-SEG 4785  
Db 267 LATKAISNDF-----KQATSNIAK--ELKSTSTELAVETKDRLSAKKET 310  
QY 4786 YSTADEL--ARTGDADLSHTSSDDDESRAQT--PSLVYLYKAGPPTGSLASKVGAAP 4841  
Db 311 FTSREAKDGKQGNQNEANAVPQESR-GTKGPGVQLQ-----KTSSTITLQAKQAP 363  
QY 4842 SVKQP---QQCEFLAAVRPPLGDLSTKDLGDPDM-DKAAVKIQAFKGVKVRKEMQOE 4896  
Db 364 KAEPTTFIRQAEDRKRTVQLMTTITQE--NPSLTGVSPSRSETEENRAGVRKSVKEK 421  
QY 4897 -----GPMFSHTGDTAEQVGDALRLLECVCVASKADYRARKWLKDGVELTGRHHIDQLG 4950  
Db 422 REPIGPQFESRPSLEAEGEIKFKSKVSGKPKDPVEMFKEGVPIKTGEGIQIYE-E 480  
QY 4951 DGTCSLLIAGLDRADAGCYTCQVSNKFGQVTHSACVVVSGSEAESSGGELDDAPRA 5010  
Db 481 DGTCHLMLKACLGSGSYSCAAENPGQTS----- 513  
QY 5011 ARRLHRLFRTPKSPAEVSDDELFLSADGPAEPPEADWQTVREDEHPICFIRFALTEARQ 5070  
Db 514 -----W-----LLTVKRP 521  
QY 5071 AVTRFQEMFATLGIGVEIKLVEQGPRRVCENISKETPAVVPPEPLSLTSDAARVFLT 5130  
Db 522 KV-----EVAPCFSS 532  
QY 5131 ELQNEVODGYPVSCVVTGQPMPSVRPKDGKLLSEDDHYMINEDQOQHQLITAVV 5190  
Db 533 VLKGTVSEGQDFVLQCVGVPVPEITWL---LNEQPIQYAHSTPEAGVAKLTVDAL 588  
QY 5191 PADMGVYRCLAENSMGVSYTKAELRLVDLTSDYDTAADATESSESYSAQGYLSRRQEGT 5250  
Db 589 PEDDGIYTCLEANNAGRASCSAQTV-----KPKSKSKAEGT 626  
QY 5251 EST-TDEGQLPQVVEELRLQVAPGTGLAKFQLKVGKYPAPRLYWFKDGQPL--TASAH 5307  
Db 627 QAAKLNKTFAPIFLKLGLTDLKVMDSQVI-MTVEVSANPCPEIILHNGKEIQETEDFHF 685  
QY 5308 RMTGKILHTLEISVTRFDSGQYAAVISNAGAAVSSARLLYRGDPDEBEKSPASDVHEQ 5367

686 EKKGNE--YSLYIQEVFPDGTGKYTCBAMNELGETQATLTVQEPDQG----- 732  
5368 LVPRMLERFPPKVKKSSITSVKVGEPVPTVHMLREEARGVLMIGDTPGYTVAS 5427  
733 -IQPWFISKRSVTAAGQVNLISCAIAGDPFFPTVHFKDQGE-----ITPQT-GEILQ 785  
5428 SAQHSLSLVDYGRHQGTCTCIASNAAGALCSALHVSGLPKVBEQEKVREALSTFL 5487  
786 NEDIFTLLRNQSRHAGQVEIQLRNQVGCSCQVSL-----MLRESSASRAEML--- 835  
5488 QGTTQALSAOQLETAADLGGQKEEPLAALKAALHLSLAETFEFLQKLSQITEMV 5547  
836 -----RDGRESA-----SGERRD-----GNYGALTGRTSGFKK----- 866  
5548 SAKITQAKLVPGGDSDEDSKTPASPRHGRSRPSSSIQSSSESESDGARGEIFDIYV 5607  
867 -----SSSETRAEEEOEDVKG-VLKERVE 890  
5608 TADVLPLCAEODAITLREGQVEVLDAAHLRLNLRVTKPKSSPSROGWSPAYLDRRLK 5667  
891 TREH-----TEE---SLAQ-QEAEQLDFRDILGKKVSTK-----SFSEEDLK 928  
5668 LSPBWGAAPPEPFGAEVSEDEYKARLSSVIQ-ELLSSEQAFVEELOFLQSHHLQHLERC 5726  
929 -----EIPAEQV---DFRANLQVQPKTLSEERKVPAPQVDFRSVLAKGT 974  
5727 PHVPI--AVAGQKAVI--FRNVRDIGRHSFLOEQOCDTDDVAMCFIKNOAFEOYL 5782  
975 PKTLPBKVPKPPKPAVDFRSVLGAKK----- 1001  
5783 EFLVGRVQAESVVVSTAIQBYKYABEALLAGDPSPQPPPLCHYLEQFVERVQVQAL 5842  
1002 -----KPPAENGSASTAPN--ARAGSEAOQATNSAPAP----- 1035  
5843 LKELIRKARNRQNCALLEQAYAVVSA--LPQRAENKLHVSLMENYPTGLEALGEPH 5900  
1036 -KPVVKEKXNDKRC--EHCAVVDGIIIGKKAENK----- 1068  
5901 HFIVWEGAPGARMKPNKHNHVFRLPHNLVICKRRDSRTDTSYVFRNMKLSIDLND 5960  
1069 -----PAASKP-----TPPPSKGT----- 1082  
5961 QVEGDDRAFEVQWQREDSVRKVYLQARTAIKSSWKEICIGIQORLALPVMRPDPBEEL 6020  
1083 -----VPPRFVNVKVRASPEVE-----GEDAQFTCTIEGAPYQIRWYKDGALLTGN 6080  
6021 ADCTAELGETVKLACRVGTGPKVISMWYKDGKAVQVDPHILIEDPGSCALLDLSLTV 6080  
1091 QDAKVADGKELVLQCRISDDPPASVSWTLDSKAIK-SSKSIVISQEGTGLSLATIEKWPE 1149  
6081 DSGQYMCFAASAAGNCSTLGLILNQ----- 6105  
1150 DGGETKCTAENAGKAEACAKVLVEDTSSSTKAAPKAEKTKKPTTLPVVLSTESSBATV 1209  
6106 -----VPPRFVNVKVRASPEVE-----GEDAQFTCTIEGAPYQIRWYKDGALLTGN 6152  
1210 KKKPAPKTPKAAATPQITQPPEDKVRAGESVELFAKVGTAFITCTWKFQKIQIGENE 1269  
6153 KFTLSEPRSGLLVLPAAASKEDLGLYECVLNRLGASARASALRIQSPMLQAOEQCHR 6212  
1270 YIKIENAEISS--KLTISSTKQEHGCGYTLVVENKLGSRQAVNL----- 1312  
6213 EQLVAAVEDTTLERADQEVTSVLKELLGPKAPGSTGLTGTGPGCPGAPALQBTGSQPP 6272  
1313 -----TVDKPD-----PPAGTFCASDIRS-----SSLTSLWYSS-- 1343  
6273 VTGTSEAPVPRVPPQLLHGGPEQEPBAIRAEQWTVPIRVEGAAMPAGCTGBELLND-V 6331  
1344 -----YDGG-----SAVOSYTVET-----WNSV 1361  
6332 HSHVRETTQTYTQAITHTAREPPSMQVITQVQAGTGTAFQAEALIEGDPQPSWTY 6391  
1362 DNKWTDLTCTSTSNVQLOLQADREYKFRVRAANVYGISSEFSQSEVVKVGE----- 1413

6392 XDSVOLVDSTRLSQOQEGTYSVLVRHVASKDAGV-YTCLAQNTGGQVLCRAELLVLOGD 6450  
1414 QOEBELKEEAEALSDEG-----KETEYVYRTVINT----- 1445  
6451 NEPDSEKSHRRKLSHFSFVKEEIGRGVFGVFRVVOHKGKILCAAKFIPILRSRTAGAY 6510  
1446 -----EOKVSDVYNIERLGSRGFGQVFRLEVEKTKGKWKAGKFFKAYSAREKENI 1495  
6511 R-ERDIALSHPLVTGLLDQPETRKTLLILELCSSEELDLRLYRKGV-VTEAEVKVYI 6568  
1496 RDEISIMNCLHHPKLVQCVDAFEKANIWMVLEWVSGGELFERIIDEDELFETERECIKYM 1555  
6569 QOLVGLHYLHSHGVHLHDIKPSNLMVHPAREDIKIDFGPAQNITPAELQFSQYSGPE 6628  
1556 ROISEGVEYIHKQGVHLDLKPENIMCVNKTGTSIKLIDFLARLEASAGSLKVLFGTPE 1615  
6629 FVSPETIQONPVSEASDIWAMGVISYLSLTCSPPAGESDRATLLNVLGRVSWSSPMAA 6688  
1616 FVAPEVINVEPIGYETDMISIGVICIILVSGLSPMGNDNETLANVTSATWDFDEAFD 1675  
6689 HLESDAKOFIKATLQAPQAPSAOCLSHPWFLKSMFAEAAHFINTQKFLRLARSWQ 6748  
1676 EISDDAKDFISNLLKDKMKRLNCTQCLQHPWLQDKTKNMEAKKLSDRMKKYMARRWQ 1735  
6749 RSLMSKYSILVMSRIPPELLRPPDSPSLGVARHLCEDTGGSSSSSSSDNELAPARAKS 6808  
1736 KTHAVRAI-----GRSSVAMISGMSGKASGSS----- 1765  
6809 LPPSPVTHSPLLHPRGFLRPSA 6830  
1766 -PTSPINADKVENEDAFLEEVA 1786  
RESULT 5  
TRIO HUMAN  
ID TRIO HUMAN STANDARD; PRT; 3038 AA.  
AC Q75922; Q13458;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Triple functional domain protein (PTPRF interacting protein).  
GN TRIO.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RX TISSUE=Brain; Tissue=Brain; PubMed=8643598;  
RA Debant A., Serra-Pages C., Seipel K., O'Brien S., Tang M., Park S.-H.,  
Streuli M.,  
RT "The multidomain protein Trio binds the LAR transmembrane tyrosine  
phosphatase, contains a protein kinase domain, and has separate rac-  
specific and rho-specific guanine nucleotide exchange factor  
domains.";  
RT Proc. Natl. Acad. Sci. U.S.A. 93:5466-5471(1996).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA Streuli M.;  
RN Submitted (SSP-1998) to the EMBL/GenBank/DBSJ databases.  
RN [3]  
RP STRUCTURE BY NMR OF 1227-1407, AND MUTAGENESIS.  
RX MEDLINE=99005194; PubMed=9790533;  
RA Liu X., Wang H., Eberstadt M., Schmichel A., Olejniczak E.T.,  
Meadows R.P., Schervanitz J.M., Janowick D.A., Harlan J.E.,  
RA Harris E.A., Staunton D.E., Resik S.W.;  
RT "NMR structure and mutagenesis of the N-terminal Dbl homology domain  
of the nucleotide exchange factor Trio.";  
RN Cell 95:269-277(1998).  
RN [4]  
RP CHARACTERIZATION.



Db 1758 DSDSAAATPQDETVEERNEGLSSGTLSSKSSSQSCGEBEE-----EGADAVPLP 1811  
Qy 5615 ---GAEOADITLREQQYVEVLDAAHPLRLWRTPTKS-SPSRQGWYSP-----AY 5661  
Db 1812 PPMALQHSLLQPSQ-----DDKASSELLVR--PTSETPSAAELVAELVKSMAL 1864  
Qy 5662 LDRRLKLSPEWGAABEF-PGE-----AVSEDEYKARLSS-----VIQELLSSEA 5707  
Db 1865 EDRPSSLLVDQDSSSPSFPNSDNLSSSSPIDEMBERKSSSLKRHHYVLQELVETERD 1924  
Qy 5708 FVEELQLOSHLQHLERCPHPTAVAGQXAVIERNVDRDGRPHSS-FLQELQOCDTD-D 5765  
Db 1925 YVRDLGVVEGYMA-LMKEDGVDDMGKQKXIVGNTHQIYDHRDFLGELEKCLDPE 1983  
Qy 5766 DVAMCFIKQAAFBQYLEFLVGRYQABSVVYSTAIOEFYKYABEALLAGDPSQPPPPPL 5825  
Db 1984 KLGLSVFVCHERRLHMYIAYCQNKPEH-IVSEVIDTFFDLKQK-----L 2028  
Qy 5826 QHYLE-----OPVERVQRYQALLKELIRKARNQNCALLQAYAVVSALPQAEKHL 5879  
Db 2029 GHRQLQTLDLLKPVQRIKXQLLKDFLYSKASLDTSELAVERAVYVNCIVPRCNDMMN 2088  
Qy 5880 VSLMENYPTLEALGEPTRQGHFTVWEGAPGARMPWKGNRHVFLFNHVLICKPRDRS 5939  
Db 2089 VGRLOQPGDKIVAQKULLQDTFLVTDQAGL-LP-RCRERITFLFEQIVIFSEPLDKK 2146  
Qy 5940 TDTV-SYVFRNMKLSIDLNDQVEDDRAFEVQWQERDSVKYLLQARTAIKSSWKE 5998  
Db 2147 GFSMPGFLFKNSIKVSCLEENVENDPCKEALTSGTGVVETILHSSSPSVRQTWIE 2206  
Qy 5999 ICGI---QORLALPWRPDPDEE-----LA 6021  
Db 2207 INQILENORFNALTSIEVQRHNSGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2266  
Qy 6022 DCTAELGETVKLACRVGTGTPKPVISWYKQKAVQVDPHH--ILIEDPDGSCALILSLTG 6079  
Db 2267 SCG---GAPSTSRSPRIPQPV-----RHHPLVLSAASQA-EADKMSG 2309  
Qy 6080 VDSGOYMCFAASAAGNCSTLCKILVOVPPRVNKV-----RASPVEGEDAQCTCTIEG 6133  
Db 2310 T-----STPGSL--PPGAAPAGSPAPRRPPGADAESE-----REA 2347  
Qy 6134 APYQIRWYKDGALLTGNKQTLSEPRSGLLVLVIRAA-----SKEDLGLY 6180  
Db 2348 EPIP-----KMKVLESFPG-----AANASGSSPDAPAKDARASLGLT 2385  
Qy 6181 ECELVNLGSRASAELEIQSPMLQAOQBCHREQLVAAVEDTTLERADQEVTVLKLKG 6240  
Db 2386 P-----LGKPRAGAASPLNSPLSAVPSLKGEPFPPS---SPLQKGSFWSST-- 2430  
Qy 6241 PKAPGPGTDLTGPG-----PCPRGA-----PALQETGSGPP- 6272  
Db 2431 PASPASRPGSTFFGDSLSLQRTFPRHAPGKDTDRMSTCSSASEQSVQSTQNGSESS 2490  
Qy 6273 -----VT-----GTSEAPVPRVP-- 6287  
Db 2491 SSNTSTMLVTHDYTAVKDEINVQGEVQVILASNQNMFLVFRATDQCPAEGWIPGF 2550  
Qy 6288 -----QPLHEGPEQEPARAOEWTPVEM---EGAAWPGAGTGL- 6331  
Db 2551 VLHTSAVIVENPD---GTLNKSSTWHTALRKXKXKXKXKXKXKXKXKXKXKX 2607  
Qy 6332 HSHVVREVTQRTYTYQAIDHTAPPPSQVTVIEVQAGTGTGTAQFEAIIIEGDPQPSVTWY 6391  
Db 2608 SNKSVKLLNPNYIDV-----PPEFVIPLEVTCTGTETVVLRCVRCGRPRASITWK 2660  
Qy 6392 KDSVQLVD-----STRLSQOQEGITYSLVLRHVASKDAGVYTCIAQNTGGQVLCABLY 6446  
Db 2661 GPEHNTLNDGHYSISYDLGEAT---LKIYGVTTEDDGYTCIAVNDMSGSSASSALRV 2717  
Qy 6447 LGGDNPEPSEKQSHRKLHSFVEKKEICRGVFGVGRVQKGNKILCAAKFIPLRGTR 6506  
Db 2718 LGPG---MDGIMVTKDNFDSYSEVAELGRGFSVVKCDQKGTGRAVATKFNKLMKR 2775

## RESULT 6

PGEM MOUSE  
ID PGEM MOUSE STANDARD; PRT; 3707 AA.  
AC Q05793; 1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Basement membrane-specific heparan sulfate  
DE protein precursor (HSPG) (Perlecan) (PLC).  
GN HSPG2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Melanoma;  
RX MEDLINE=92078153; PubMed=1744087;  
RA Noonan D.M., Fullie A., Valente P., Cai S., Horigan E., Sasaki M.,  
RA Yamada Y., Hassell J.R.;  
RT "The complete sequence of perlecan, a basement membrane heparan  
RT sulfate proteoglycan, reveals extensive similarity with laminin A  
RT chain, low density lipoprotein-receptor, and the neural cell adhesion  
RT molecule.";  
RL J. Biol. Chem. 266:22939-22947(1991).  
RN [2]  
RP SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=89034110; PubMed=2972708;  
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M.,  
RA Yamada Y., Hassell J.R.;  
RT "Identification of cDNA clones encoding different domains of the  
RT basement membrane heparan sulfate proteoglycan.";  
RL J. Biol. Chem. 263:16379-16387(1988).  
CC -|- FUNCTION: This protein is an integral component of basement  
CC membranes. It is responsible for the fixed negative electrostatic  
CC charge and is involved in the charge-selective ultrafiltration  
CC properties. It serves as an attachment substrate for cells.  
CC -|- SUBUNIT: Purified perlecan has a strong tendency to aggregate in  
CC dimers or stellate structures. It interacts with other basement  
CC membrane components such as laminin, prolargin and collagen type  
CC IV.  
CC -|- SUBCELLULAR LOCATION: Extracellular.  
CC -|- TISSUE SPECIFICITY: Found in the basement membranes.  
CC -|- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS, AS WELL AS N-LINKED  
CC AND O-LINKED OLIGOSACCHARIDES.  
CC -|- SIMILARITY: Contains 4 LDL-receptor class A domains.  
CC -|- SIMILARITY: Contains 11 laminin EGF-like domains.  
CC -|- SIMILARITY: Contains 3 laminin IV domains.  
CC -|- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.  
CC -|- SIMILARITY: Contains 3 laminin G-like domains.  
CC -|- SIMILARITY: Contains 1 EGF-like domain.





FT DISULFID 1886 1932 BY SIMILARITY.  
FT DISULFID 1976 2021 BY SIMILARITY.  
FT DISULFID 2073 2118 BY SIMILARITY.  
FT DISULFID 2170 2215 BY SIMILARITY.  
FT DISULFID 2268 2313 BY SIMILARITY.

Query Match 1.9%; Score 767; DB 1; Length 3707;  
Best Local Similarity 20.8%; Pred. No. 2.2e-18;  
Matches 870; Conservative 464; Mismatches 1543; Indels 1308; Gaps 218;

QY 1193 PTEVTWYKDGKLSKSSKVRMEYKGTTRRLVVOQVGKADAGEYSCBAGQORVSFOLHIT 1252  
DB 98 FOLEDASAKREFVESEAVVEKLE-----PEYRKIPGQIVS----- 133

QY 1253 EPKAVFAKE-----OLVHNEVTEAGASATLSCEVAQAQTEVTWYKDGKLSKSSKVRIEA 1308  
DB 134 ---VVFKELGWVFEVLVDGSEGNADGSQIQEV--LHTVVSSGSIOPYVTSPNGFKFR 188

QY 1309 AGCMRQ-----LVVQOAGQADAGEYTC--AGGORLSFHLDVSEPKAVFAKEQLAH 1357  
DB 189 LGTVPPQPRVCTETEFACHSYNECVALEYRCRRPDCRDMDELCEPVPPELSSSTPA- 247

QY 1358 RKVOAEAGAIATLSCEVAQAQTEVTWYKDGKLSKSSKVRMEAVGCTRRRLVVOQAQOAT 1417  
DB 248 ---VGKVSPLPLWPEAATPPPPVTHGPQFLPLSPVPGPSACG-----P- 287

QY 1418 GEYSCEAGQORLSFSLDVAEPKVFVFAKEOPVHREVOAQAGASTTSLCEVAQAQTE----- 1472  
DB 288 QEASCHSH-----CIFRDYLCQOECRGSDELGCASPPCEPNEFACENGHCA 338

QY 1473 -VMYKDGKLSKSSKVRMEAVGCTRRRLVVOQAQO-ADAGEYSCBAGQORLSFHLHVAE- 1529  
DB 339 LKLWRCDG--DFDCEDRTDEANGS---VKQPEGVCGPTHFQCVSTNRCPASFHCDEE 391

QY 1530 ---PKAVFAKEQASREVQAEAGTSATLSCEVAQAQTE-VTWYKDGKKL 1574  
DB 392 SDCPRDSDFGCMPPQV---TPQOSIQASRGQTVTFCATGVPTPIINMLNWGHI 447

QY 1575 SSSSKVRMEAVGCTRRRLVVOBAGADAGEYCKAGDOR-LSFHLHVAEPKVFVFAKEQPAH 1633  
DB 448 PAHPRVTWTSSEGRGTLIIRDVKEADQAGYTCAMNSRGMVFGIPDGVLVLPVQGPCCPD 507

QY 1634 REVQAEAGASATLSC-----EVAQAQTEVTWYKDGKLS- 1667  
DB 508 GHFYLEDSASC-LPCFCFGVTVNVCSSLR---FRDQIRLSFDQPNDFKGVNVTWPSQGV 563

QY 1668 ---SSSKVRV-----EAVGCTRRRLVVOQA-----GOADA-----GEYSCE 1699  
DB 564 PPLSSTQLOIPALQEFOLVLSRRFLVHDAFWALPKQFLGNKVDSYGFRLYKRVYELA 623

QY 1700 AG-----GQRLSFRLL-----VAELEPQISE-----RPCR-EP 1728  
DB 624 RGMLEPVQKPDVILVAGYRLHSRGHTPTHPTGLNQVQLSEBHHVHESGRPVQRAEML 683

QY 1729 VKHEHDIILATLATPSAATVWVKDGEVIRRSKRHETASQGDTHLTLVHGAQVLDASAI 1788  
DB 684 QALASLEAVLLQTVNTKXSV-GLSDIV-----MDTTVTHT-TLHGRA---HSV 728

QY 1789 YSCR--VGAEGQDFPVQVEEAAKFCRL-----LEPV----- 1818  
DB 729 BEECRCPIYSG---LSCSCDAHPTRVPGPGPYLGTCSGNCNGHASSCDPVYGHCLNCQ 784

QY 1819 -----C-----GELGTVILACELSP-----ACAE 1838  
DB 785 HNTGPOQCKKPGFPGDATKATATACRPPCPVIDASRRFSDFCTFLDTGQATCDACAP 844

QY 1839 VVWRCGNTQPRVKRFQVMAE3-----PV-----RSLTVLGLRAEDAGEY-----VCESR 1883  
DB 845 -----GYT-----GRCECAPGYEGNPIQPGKCRPTTQEIVRCDERGLTSETCRCK 895

QY 1884 DDHTSAQITVSVPRVVKPMGSLSTVVAEEGEATFQCV---VSPSDVAVVWFRD---GAL 1937  
DB 896 NN-----VYGRLCNECDSGSEFHLKQNPDGCLKFCMGVSRQCSSSSMSRAQVLGAS 947

QY 1938 LOPSEKFAISQSGASHSLTISDLVLEDA-GQIT-----VBA 1972  
DB 948 EQPSQ-FSLSNAAGTH--TTSEGVSPAPGELSFSSPHNLLSEPYFWSLPASFRGDKVTS 1004

QY 1973 EGASSSAALRVREAPVLFKKLEBPQTVBERSVTLLEVELTR-----PWPELFW 2020  
DB 1005 YGELAFRTVQPRSSAPLHQPLVVLQGNVILHSHASRDPSPQPSNFIVPQEQAM 1064

QY 2021 TRNATALAPKKNVEIHAEQARHLVLHNVFADRGFFGCETPDDTKOAKLTVMEMQVRLV 2080  
DB 1065 QRPDQGPATREHLLMALAGIDALLI-----QASYT----- 1094

QY 2081 RGLQAVEAREQGTATMEVOLSHADVDGSMTRDGLRFQOQPTCHLAVRGMHTLTLGLRP 2140  
DB 1095 ---QCPAESLSG-ISMVA-----VPENTGQDSAREVEOCTCPPVGRGSDCDDTGYTR 1146

QY 2141 EDSGLMVFKAEGVHTSARLVVTPVLPVSFRPLQDVVTEKEKVTLECELSRNVDRWLK 2200  
DB 1147 VPSGLVGTGCRCNCHGHSETCEPETGACQCOH--HTE---GASCEQCQPGY----- 1194

QY 2201 DGVLELAGKTMALAAQGA---CRSLTIYRCEPADQGVYVC-----DAHDAQSSASVKVQGR 2253  
DB 1195 ---YGDQORGTQDQCPQPCYGAAGAQAHTCFDLDGHTPCTDCSCPSGHSR 1244

QY 2254 TVTLI-----YRRVL---AEDAGEIQFVAENA-ESRAQLRVKELFV 2290  
DB 1245 HCERCAPGYGNPSQCPCHRDQVPEVLGCGDPHGSISSOCDAAGQCCRAQVEGRSC 1304

QY 2291 TLVRPLRDKIAEKHGVLEC-----QVRSASAQVRW-----FKG-----S 2326  
DB 1305 SHCRPHHFLSANSPEGLPCFCMGVTQCCASSSRQLISTHFAPGDFQGFALVNPORN 1364

QY 2327 QELQCPKVELVDGLYRKLIISD-VHAEDETY---TCDAGDVKTSQFFVEEOSIT 2381  
DB 1365 SQTGTGFTVEPHDG--ARLSFNFAHLCQESFYQLPEIYQGD-KVAYAGKRLVTL 1421

QY 2382 VRGLQDVTVMEPAPAFECETSIPIVVRPKWLLGKTVLQAGNVGLQEGETVHRLMLRRT 2441  
DB 1422 TAGPQGSPLDPP-----IQITGN-----NIMLVAS 1447

QY 2442 CSTMTGCVHFTVKGSRSSARLVVDIPVVLTPLEPKTGRE-----LQSVLSCDFR 2493  
DB 1448 QALQOP-----ERRSYEIIFRE--EFWRPPDQGPATREHLLMALADDELLVRATFS 1498

QY 2494 PAFKAVQWYKDDTLPSPSEKFKMSLEG-----QMAELRILIMP----- 2532  
DB 1499 SVPERA-----ASISAVSLEGAQPGSPSGPRALEVEECRPPGVGLSCQDCAP 1546

QY 2533 ---ADAGYV-----RCQAGSAHSSTEVTVAREVTVTGPL 2564  
DB 1547 GYTRTSGSLYLOQCELCNCGHSDLCHPETGACSRQHNTAG-----EFCELCATGY 1599

QY 2565 QAAEA-TEEGWASFSCELSHEDBEVWSLNGM-----PLYNDSFHSIHKGRH 2612  
DB 1600 GDATACTPDDCQPCACPLTNPENMFSRTCESLGAGYRCTACEPGYTGQYCEQCAPG--- 1656

QY 2613 TLVLKSIQADAGIVRASSLKVSTSRLEVRKP---VVLKALDDLSAEERGTALQCE 2669  
DB 1657 ---VEGDPN-VQGGRCQPLTKESLEVOIHPSSRVV-----POGGPHSLRQ 1698

QY 2670 VS-DPEAHVW-RKDGVLGSDKYDFLHTAGTRGLVWH--DVSPEDAGLYTCH---VG 2721  
DB 1699 VSGSPHYFYWRREDRPLPSSAQORH-----QSEHLPSPVQSDAGVYICTENLIH 1752

QY 2722 SEETRARVRVD-----LHVGITKRLKTMVELEGSCSEFVLSHESASDPAMWT--VGG 2774  
DB 1753 TNSRAELLVAEAPSKPIWTVVEEQ-RSQSVRPGADVTFICTAKSKSPAYTLVWTRLHG 1811

QY 2775 KTVGSSSRQATQGRKYLIVVREAPSDAGEVWFSVRL-----TSKASLIYR-----BR 2825  
DB 1812 KLPFSRAMDFNG-----ILTRNVQPSDAGTYVCTGSGNFMAMDQGTATLHVQVSGTSTA 1864

QY 2826 PAATIKPLEDOWAPGEDVELCELSSRAGTP-VHML--KDRKAIKRSOKYDVVCEGTMM 2882  
 Db 1865 PVASIH-POLITVPGQOAFRCBATGNPTMLEWIGPGSQGQPAKAQIHN-----GI 1916  
 QY 2883 LVIRGASLKADAGEYTCVEASKSTASLHVEE-----RANCFTTELTNLQVEEKG 2931  
 Db 1917 LRLPAIBSDQGYLCR---ALSSAGQHVARAMLQVHGSGPRVQVSPE---RTQVHEGR 1970  
 QY 2932 TAVTCKTEH-PAATVTRK--GELLEIRASQKOPSOEGTLTLTISALEKASDIYTCD 2988  
 Db 1971 TVRLYCAAAGVPSASITWRKEGSLPFR---HOAHGS-----RLRLHMSVADSGEYVC- 2021  
 QY 2989 IGQAQSAQLLVQRRVHIIEBLEDVDVQEGSSATFCRI-SPANYEPVHVFLDKTLPAN 3048  
 Db 2022 -----RAN-----NNIDAQETS-----IMISVSPSTNSP-----PAPAS 2050  
 QY 3049 ELNEIDAQPGYHVLTURLQALDKSGTIFYEAGQORASALRVTKEPSVFSRELTATIT 3108  
 Db 2051 -----PAPIRIESSSRVA 2064  
 QY 3109 EGEDTLVCESTCDIP-----MCWTKDGKTLRGSARCQLSHEGHAQLLITGATLQDS 3162  
 Db 2065 EGQTLDL-----NCVPGHAAQVTHWKRGS-----PTHQTHGSRRLRQVSSADS 2113  
 QY 3163 GRYKC-----EAGGACSSIV-----RVH-----ARPRFQEAALKOLEVLEGGAAATLRCV 3207  
 Db 2114 GEYVCSVLSSGPLEASVLVSITPAAANVHIPGVVPPIRIETS--SSRVAEGQTLDLSCV 2171  
 QY 3208 LSSVA-APVWCYGNVLRPGDKYSLRQBGMLLVNRLRPODSGEYSC-----SEGDQT 3262  
 Db 2172 VPGQAHAQVTHWKRGSGLPAGHV-----HGHMURL--NRVSPADSGEYSCQVTGSSGTLE 2225  
 QY 3263 TSATLTIVA-----LPAQFIG-----KLRNKEATEGATATLRCGL--SKTAPVEMWRKGE 3310  
 Db 2226 ASVLVTEASRPSIPAPGLAQPVIYESSSHLTGQTVLKVVPQQAHAQVTHWKRGS 2285  
 QY 3311 TLRGDRYCLQDQAMCELQIRGLAMVDAABYSC-VCG-----BERTSASLTIRPMAHFI 3365  
 Db 2286 SLP-----ARHOTHGSLLELYQLSPADSGEYVCQVAGSSHPHEASFCLTV-PSQNS 2338  
 QY 3366 GRLRHOE-SIB-----GATATLRCELSKAA--PVEWRKGRSLRDCDRSLRQDGA 3413  
 Db 2339 FLRSPTISIEPPSTVQOGDASFCLIHGAMPKIVEMKIRDOELEDNVH--ISPNGS 2396  
 QY 3414 VCELOICGLAVD--AGEYSCVCEBERTSATLTWKALPAKTEGLRNEEAVE---GATAM 3468  
 Db 2397 IITIVAPGATMEPTACVASNVYMAQSVNLSVHGPP--TVSVLPEGPHVYKMGKDIT 2453  
 QY 3469 LWCELSKVPVEMWRKGPENLRDGRYTLRQE-----GTRCELOICGLAMADAGEYLC--- 3520  
 Db 2454 LEC-----ISSGEPSRPTWGLGIPVLEBPMGLMNSHMLKIASVKPSDAGTYVCOAQ 2509  
 QY 3521 -VCQERTSATLTI---RALPARFIEDVKNQE---AREGATAVLQCELSNAAP--VEWRKG 3572  
 Db 2510 NALGTAQXQVELIVDTGTVAPTQVQVBESELTLEAGHTATLHCSATGNPPPTIHWK- 2568  
 QY 3573 SETLRDGRYSLRODGYKCELOIRGLAMADTGEYSC-----VCQERTSAMLTRALPIKF 3628  
 Db 2569 ---LRAPLFWOHRIEGN--TLVTPRVACQDSQGVYCNATNSAGHTEATVVLHVESPPYA- 2622  
 QY 3629 TEGLRNBEATGATAVLRCESLKAAPVEM-WKHETLRDQ--DRHSRODGAARCELQIR 3684  
 Db 2623 TIIPHTSAQGNLVQLQCLAHGTPPLTYQW-----SLVGGVLEKAVVRNQ-----LLRLE 2674  
 QY 3685 GLVABDAGEYLCMGCKERTSAMLTRAMPKSKFIEGLRNEEATEGDTATLW----- 3734  
 Db 2675 PTVPEDSGRYCQVSNRNVGSAEPAQVLVQSSNNLPTDTSIPGCSPTPTVQVTPQLETRNI 2734  
 QY 3735 ---CELSKAAPVE-----WRKHETLRDGRHSRLRQDGRSCELBQIRGLAVVDAGYSCV 3785  
 Db 2735 GASVEFHCAVPNERGTHRLWLKEGGLPPG--HSV-QDG--VLRIONLDQCCGYIVCQ 2788  
 QY 3786 C-----GOERTSATLTVRALPARFIEDVKN--QEARAGATAVLQOC-ELSKAAP-VEWRKGS 3837

Db 2789 AHGPHGQAATAQLIVQALPSVLI-NVUTSVHVVVHSHVEFECLALGDPKQVWTMSKV 2847  
 QY 3838 EFLRGDRYSRQDQTRCELOIRHLSVADTGEYSC-----VCQERTSATLTVRAPQPVFR 3893  
 Db 2848 GHLRPG-----IVQSGT--IIRIAHVELADAGYRCATNAAGTTQSHVLLVQALPOIST 2901  
 QY 3894 EPLQSLQAEESTATLQCELS-EPTATVWSKGGIQLQANGRRBRRLQCGCTAEIQLVLOLQ 3952  
 Db 2902 PP--BIRVPAGAAVPCWAGPYTPALTWSKVDGLDPPDSRLNNM-----LMLFSVR 2993  
 QY 3953 REDTEYCTC-----GSOATSATLTV-----TAAPVRFRLRLOHQEVDGEGTAHLCC 4000  
 Db 2954 PEDAGTYVCTATNRGKVKAFAYLQVPERVIPYPTQTPSYSLPLPTIKDA----- 3003  
 QY 4001 ELSRAGASVEWRKGSLOLFPCKAKYQMVQDGAABLLVRGVEQEDAG----- 4046  
 Db 3004 -----YRKFEI-----KITFRPDSADGMLLYNGQKRSPTNLNLRQPDFISFGLV 3047  
 QY 4047 ---DYTCDTGTOSMASLSVRPERP---KFK--TRLQSELET---GDIARLCCQLSD 4093  
 Db 3048 GRPPEFRDAG--SCMA--TIRHPTPLAGQFHTVTLRLSLTQGLSIVGNLAPV---N 3098  
 QY 4094 AESGAVVQWLKEGVELHAG--PKY-----EMRSQATRELLIH----- 4129  
 Db 3099 GTSQKFGQLDLNEELYLGYPDYGAIPKAGLSGFGVCVRELRIQG--BEIVFHDVNL 3156  
 QY 4130 -----OLEAKDTGEYACVTCGQKTAASLRVTEPEVTVIRGLVDAE- 4169  
 Db 3157 THGISHCTCDRPPCQNGGQDQDSESSSYTCVPAGFTAAAVNIRKCTATPSLWADATC 3216  
 QY 4170 VTADEDVEFSCEVRAGATGYQWCLQGLPLOSNEVTEVAVRDRGRIHTLRKLVGTPEDAGT 4229  
 Db 3217 VNRPDGRGYTCR--CHLGRSGVR--CEEGTV-----TTPSMSGA 3252  
 QY 4230 VSF-----HLGHASSAQLTVRAPEVITILEPQVQLSEGQDASFQ--CRLSRASSQ--EAR 4282  
 Db 3253 GSYLALPALTUTHHELRLDV---EFKLEPNGLILFSGKSGPVEDFVSLAMVGGHLEFR 3309  
 QY 4283 WALGG-----VPLQANEMNDITVEQ-----GTLHL-----LTLH--- 4311  
 Db 3310 YELGSLAVLRSHBPLALGRHVRSAERLNKDSLVRDGGPRVLRSPGKSGOGLNLHTLL 3369  
 QY 4312 -----KVTLEDAGTVSFHVGTCSSEAOQLKVTAKNTVVRGLENVVALEGEALFECQLS 4364  
 Db 3370 YLGGVEPSVQSPATNMSAHFHGCVGE--VSVNGKR-----LDLTYSLGSGVGGCYDS 3422  
 QY 4365 OPEVAHAHTWLLDDPVRTSEN-----ABVVF---FENGLRHLHLLKLRP---QDSRVTF 4414  
 Db 3423 SP-----CERQPCRNAGTCMPAGEYEFQCLQDGFKGDLCEHENPCQLHBPCC--- 3470  
 QY 4415 LAGDMVTSAFITVRGWRLEIIEPLKNAAVRAGAQAARTCTLSEAVPVE-----A 4464  
 Db 3471 LMGG-----TCRGARCLCPGFSGRCCQG--AGYGVWESDWHPEGSGNDAPGOYGA 3521  
 QY 4465 SWYINGAAVQPDSS-----DMVTADGSHQALLLRSAQPHHA----- 4501  
 Db 3522 VFYDNGFGLCPNFSRSLPEVPETIPEVITSTADG-----LJLWQGVVREASRKFIS 3577  
 QY 4502 -----GEVTFACRDASARATVLGLPDPPEDAEVAHSHHTVTLSWAAPSMDGGGCLG 4566  
 Db 3578 LCLQDCHLVFVQLGSGEARL-VSG--DPINDGE-----WHRIT-----ALREGQSG--S 3622  
 QY 4557 YRVEKEGATGWRILCHELVPGPECVVD-----GLAPGETYFRVAAV--GPVGAPEP 4607  
 Db 3623 IQVGEDILVUGR-----SPGNVAVNTKDIIVYGAPD-----VATLRGRFSSGIT 3669  
 QY 4608 VHLPTQVRLAEPKPPVPPQPSAPESRQVAGEDVSLLEVVAEAG 4652  
 Db 3670 GCIKNLVLTARPAPPPPP-----LDLQHRQAQ 3699

KWLS BOVIN  
ID - KMLS\_BOVIN STANDARD; PRT; 1176 AA.  
AC Q28824;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Myosin light chain kinase, smooth muscle (EC 2.7.1.117) (MLCK)  
DE [Contains: Telokin].  
GN MYLK.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Stomach;  
RX MEDLINE=93203148; PubMed=1284247;  
RA Kobayashi H., Inoue A., Mikawa T., Kuwayama H., Hotta Y., Masaki T.,  
RA Ebashi S.  
RT "Isolation of cDNA for bovine stomach 155 kDa protein exhibiting  
myosin light chain kinase activity.";  
RL J. Biochem. 112:786-791(1992).  
CC -!- FUNCTION: CALCIUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR  
SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SERINE  
IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT  
FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS.  
CC -!- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin  
light-chain] phosphate.  
CC -!- SUBUNIT: ISOFORM TELOKIN BINDS CALMODULIN.  
CC -!- ALTERNATIVE PRODUCTS:  
Event-Alternative initiation;  
Comment=At least 3 isoforms, Non-muscle, Smooth-muscle (shown  
here) and Telokin, are produced by alternative initiation.  
CC Isoform Non-muscle is the longest and telokin is a C-terminal  
section without catalytic activity;  
CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.  
CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; S57131; AB25794.1; -  
CC PIR; JN0583; JN0583.  
CC HSSP; P56276; 1TLK.  
DR InterPro; IPR008957; FN III-like.  
DR InterPro; IPR003961; FN III.  
DR InterPro; IPR003962; FNIII subd.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003598; IG\_c2.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR008271; Ser\_thr\_pkin AS.  
DR Pfam; PF00041; fn3; 1.  
DR Pfam; PF00047; ig; 2.  
DR Pfam; PF0069; pkinase; 1.  
DR PRINTS; PR00014; FNTYPEIII.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00060; FN3; 1.  
DR SMART; SM00408; IGC2; 2.  
DR SMART; SM00220; S\_TKc; 1.  
DR PROSITE; PS50835; IG LIKE; 3.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR Transferrase; Serine/threonine-protein kinase; Calmodulin-binding;  
KW ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;

KW Alternative initiation.  
FT CHAIN 1 1176  
FT CHAIN 1022 1176  
FT CHAIN 1022 1022  
FT INIT MET 100 291  
FT DOMAIN 100 111  
FT REPEAT 112 123  
FT REPEAT 124 135  
FT REPEAT 136 147  
FT REPEAT 148 159  
FT REPEAT 160 171  
FT REPEAT 172 183  
FT REPEAT 184 195  
FT REPEAT 196 207  
FT REPEAT 208 219  
FT REPEAT 220 231  
FT REPEAT 232 243  
FT REPEAT 244 255  
FT REPEAT 256 267  
FT REPEAT 268 279  
FT REPEAT 280 291  
FT DOMAIN 356 444  
FT DOMAIN 498 586  
FT DOMAIN 603 673  
FT DOMAIN 725 980  
FT DOMAIN 972 1035  
FT DOMAIN 1069 1158  
FT NF\_BIND 731 739  
FT BINDING 754 754  
FT ACT\_SITE 846 846  
FT DOMAIN 1171 1176  
SQ SEQUENCE 1176 AA; 128824 MW; P53DC6D4D42D4B97 CRC64;  
Query Match 1.6%; Score 645.5; DB 1; Length 1176;  
Best Local Similarity 23.9%; Pred. No. 78-15;  
Matches 137; Conservative 133; Mismatches 292; Indels 203; Gaps 20;  
QY 6014 PPFEEELADCTAELGETVKLACRVGTGTPKPVISWYKDGKAVQVDPHHILIEDPGS-CAL 6072  
DB 356 PPFEEKLQDLHVAEGKLLQCRVSDPDPTATITWLTNGKTLTKTFIVLSQ--EGSLCSV 413  
QY 6073 LIDSLTGVDSQVMCPAASAGNCSTLGLKILVQVP--PFFVN-----KVR-----ASPV 6120  
DB 414 SIEKALPFEDRLGKVCVAKNSAGQAESSCQVTVDPDPTSENAKAPEMKARRPKSLPPV 473  
QY 6121 EGEDAQFTCTIGAPY-----POIRWYKDGALLTGN-----K 6153  
DB 474 LGTESDATVKKPAPKTPPKAAMPQIIQFPDQKVRAGESVELFGKVAGTQPICTWVK 533  
QY 6154 FQTLSEPRSGLLV-----LVIRAASKEDLGLYECELNRLIGSARASAEIRIQSPMLQ 6205  
DB 534 FRKQIQDSEHIKVENSEQSGSKLTIRAAQEHGCGCYTLLENKLGSRQAQVNL----- 585  
QY 6206 AQEQCHREQLVAAVEDTTLERADQEVTSVLKLLGPKAPGPTGDLTGPQCPRGAPALQ 6265  
DB 586 -----TVVDKFPD-----PAGTFCAS 601  
QY 6266 ETGSQPPVGTSEAPAVPRVPFQPLLHEGPEOPEAIAPAEQWTVPIRMGAAPGAGTG 6325  
DB 602 DIRSS-----SLTSWYSSYDGSVA 623  
QY 6326 EL-----LNDVHSHVRE--TTORTYTYQATDHTARPPSMQVTTIEDVQAOTGGTAQFEAL 6380  
DB 624 QSYSVEINSDVKTKWELATCRSTSNVQDILLPDREYKFRVRAINIVGTSEPSQSELT 683  
QY 6381 EGD-POPSVTWYKDSVOLVDSTRLSQOQGGTYSVLRLRHVASKDAGVYTCLAQNTGGQVL 6439  
DB 684 LGKPREEP---KDEVEVSDDEKEPEVD-----YRTVTVT----- 717  
QY 6440 CKAEILLVLDGNDPEPDKSEKSHRKLKHSFYVEVEKEICRGVFGFKVRVQHKGNKILCAKFI 6499



```
Qy 6073 ILDSLTGVDGQYKCFASAGNCSTLGLKILVQPPRFVN-----KVR-----ASPFVFG 6122
Db 387 SIEKALPEDRLGLYKCAKNAEAECSCHVTYVHDAPASENAKAPKMSRRPKSSLPFVLG 446
Qy 6123 EDAQFTCTIEGAPY-----POIRWYKDGALLTGNKFTL----- 6157
Db 447 TESDAIVKXPAKTPPKAATPPQIPQFPEDQKVRAGERVELRGKVAGTQPICTWMKPR 506
Qy 6158 -----SEPSGLLVLVIRAAKEDGLYECVLNRLGSRASAEIRIOSPML 6204
Db 507 KQIQDSBHKVENSEAGSKLTIL---AARGHCGCYTLLVENKLGSRQAQVNI----- 556
Qy 6205 QAQOQCHREQLVAAVEOTTLERADOEVTSLKRLGLPKAPGPGSTGDLTGPGCPRGAPAL 6264
Db 557 -----TVVDKDDP----- 6270
Qy 6265 QETGSPPTVTGTEAPAVPRVPOPLLHEGPEOEPEAIARAEQWTVVIRMEGAAMPAGT 6324
Db 572 SDIRSS-----SULTSYGWSYDGGSA 593
Qy 6325 GELLWDVHSHVRETTORTTYQAID---THTAEPSPMQVTIEDVQAQTGGTAQFEAL-I 6380
Db 594 -----VQSYSE-----IWDSDVKWTELACRSTSNVRDLDPDRYKPRVRAINV 640
Qy 6381 EGDQPSVTYKDSVQLVDSTLSQLQOEGTYSILVRHVASKDAGVYTCIAQNTGGQVLC 6440
Db 641 YGTSEP-----SQESELIT-----VGEKPEEP----- 662
Qy 6441 KAEILLVGGNEPDSKQ-----SHRRKLHSFVEKVEEIGRGVFGKVEVQKGNKILC 6494
Db 663 KDEVEVSDDDKEPEVDYRTVTNTQKVSDFYDIEERLGSGKFGQVFLVKKTKGIW 722
Qy 6495 AAKIPLRSRRPAAQYERD-----ILAAASHPLVTGLDDQFETKTLILILELCSSE 6547
Db 723 AGKEF-----KAYSACEKENIPAEIGIMNCLHHPKLVCQVDAFEERKANIMVLEIVSG 776
Qy 6548 ELLRLYRKGV-VTEAEVKVYIQQLVEGLHVLHSHGVHLIDKPSNLMVHPAREDIKIC 6606
Db 777 ELFERIIDEDELTRECIKMYRQISEGVYIHRQGIWHLDLKPENIMCVNKTGTRIKLI 836
Qy 6607 DFGQANITPAELQFQSGSPFVYSBETIQONPVSEASDIWAMGVSYLSITCGSPPAGE 6666
Db 837 DFLARRLENAGSKVLFGTPEFVAPEVINTPEISYATDMWSIGVICYILVSGLSPMGD 896
Qy 6667 SDRATLNLVGRYSWSPMAHLSEDAKDFIKATLQAPQAPSAQAQCLSHFWFLKSMF 6726
Db 897 NDNETLANVTSATWDFDEAFDEISDDAKDFISNLLKDKMKNLDDCTQCLQHPWLMKDTK 956
Qy 6727 ABEAHFINTKOLKELLARSWRQSLMSYKSLVWRSIPELLRGPDPSPSLGVARHLCDT 6786
Db 957 NNEAKKSKDRMKYMARRWKQKTGNVRAIGRLSSWAMI----- 996
Qy 6787 GGSSSSSSSDNLAPFARAKSLPSPVTHSPHLLPRGFLRPSASLPEEAESERSTEAP 6846
Db 997 SGLSGRKSSTGS-----PISPLT-----AERLETEEDVSQAFLAV 1032
Qy 6847 A---PPASP 6852
Db 1033 AEEKPHVKP 1041

RESULT 9
MYPC_MOUSE
AC 070468; 088997; STANDARD; PRT; 1270 AA.
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myosin-binding protein C, cardiac-type (Cardiac MyBP-C) (C-protein,
DE cardiac muscle isoform).
GN MYBPC3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FVB; TISSUE=Heart muscle;
RA Yang Q., Sanbe A., Osinska H., Hewett T.E., Kelvitsky R.,
RA Robbins J.;
RT "A mouse model of human familial hypertrophic cardiomyopathy.";
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart muscle;
RA McDonald K.S., Hollander M.S., Moss R.L.;
RT "Sequence of the cardiac isoform of murine myosin binding protein-C
RT (MyBP-C) cDNA";
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: THICK FILAMENT-ASSOCIATED PROTEIN LOCATED IN THE
CC CROSSBRIDGE REGION OF VERTEBRATE STRIATED MUSCLE A BANDS. IN VITRO
CC IT BINDS MHC, F-ACTIN AND NATIVE THIN FILAMENTS, AND MODIFIES THE
CC ACTIVITY OF ACTIN-ACTIVATED MYOSIN ATPASE. IT MAY MODULATE MUSCLE
CC CONTRACTION OR MAY PLAY A MORE STRUCTURAL ROLE.
CC -!- PTM: SUBSTRATE FOR PHOSPHORYLATION BY PKA AND PKC. REVERSIBLE
CC PHOSPHORYLATION APPEARS TO MODULATE CONTRACTION (BY SIMILARITY).
CC -!- SIMILARITY: belongs to the immunoglobulin superfamily. MyBP
CC family.
CC -!- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 3 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF059576; AAC14570.1; -.
DR EMBL; AF097333; AAC64202.1; -.
DR HSSP; P56276; 1TLX.
DR MGD; MGI:102844; Mybpc3.
DR GO; GO:0005936; P:muscle contraction; IPI.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR003951; FN_III.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00041; fn3; 3.
DR Pfam; PF00047; ig; 8.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00408; IGc2; 1.
DR PROSITE; PS50835; IG_LIKE; 7.
DR Immunoglobulin domain; Actin-binding; Cell adhesion; Muscle protein;
DR Thick filament; Repeat; Phosphorylation.
KW DOMAIN 100 150
FT DOMAIN 151 254 PRO-RICH.
FT DOMAIN 358 448 IG-LIKE C2-TYPE 1.
FT DOMAIN 449 539 IG-LIKE C2-TYPE 2.
FT DOMAIN 540 629 IG-LIKE C2-TYPE 3.
FT DOMAIN 641 767 IG-LIKE C2-TYPE 4.
FT DOMAIN 768 865 IG-LIKE C2-TYPE 5.
FT DOMAIN 866 966 FIBRONECTIN TYPE-III 1.
FT DOMAIN 967 1061 FIBRONECTIN TYPE-III 2.
FT DOMAIN 1062 1176 FIBRONECTIN TYPE-III 3.
FT DOMAIN 1177 1270 IG-LIKE C2-TYPE 7.
FT MOD_RES 273 273 PHOSPHORYLATION (BY PKA AND PKC) (BY
FT SIMILARITY).
FT MOD_RES 282 282 PHOSPHORYLATION (BY PKA AND PKC) (BY
FT SIMILARITY).
FT MOD_RES 302 302 PHOSPHORYLATION (BY PKA AND PKC) (BY
FT SIMILARITY).
FT CONFLICT 32 32 E -> G (IN REF. 2).
FT CONFLICT 39 39 M -> K (IN REF. 2).
FT CONFLICT 113 113 E -> K (IN REF. 2).
FT CONFLICT 249 249 N -> S (IN REF. 2).
```





FT DISULFID 1107 1149 BY SIMILARITY.  
 FT DISULFID 1193 1241 BY SIMILARITY.  
 FT DISULFID 1281 1324 BY SIMILARITY.  
 FT DISULFID 1367 1425 BY SIMILARITY.  
 FT DISULFID 1465 1511 BY SIMILARITY.  
 FT DISULFID 1554 1613 BY SIMILARITY.  
 FT CARBOHYD 159 159 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 265 265 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 339 339 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 697 697 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 726 726 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 730 730 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 741 741 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 886 886 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1104 1104 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1138 1138 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1251 1251 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1462 1462 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1476 1476 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT VARSPLIC 1632 1709 OSMGSEVMAFOKETTQIDPDAACTETSTCAPLG ->  
 FT FT GEGGLHGHSAKPS (in isoform 2).  
 FT FT /FTID=VSP\_002571.  
 FT VARSPLIC 1666 1709 RRRRVKQSGMSENVMAFOKETTQIDPDAACTETSTCAP  
 FT FT PLG -> SSILMQPHVRPQVPHPWADQWCCPLSGGESQ  
 FT FT NL (in isoform 3).  
 FT VARIANT 239 239 /FTID=VSP\_002572.  
 FT FT K -> R (in dbSNP:625372).  
 FT VARIANT 919 919 /FTID=VAR\_014136.  
 FT FT H -> P (in dbSNP:709012).  
 Query Match 1.4%; Score 572.5; DB 1; Length 1709;  
 Best Local Similarity 23.0%; Pred. No. 3.2e-12;  
 Matches 399; Conservative 187; Mismatches 650; Indels 501; Gaps 92;  
 QY 2703 LVHDSVPDAGLYTCHVGSEETEARVRHDLHVGITK--RLKT-----NEVLEGSCSFE 2756  
 DB 100 LLLKDFPDSGNSNFFETSEVNRSDVKGTLVTVEEPVETIASPVELLEGTEVDNF 159  
 QY 2757 C-----VLSHESADPAMTVGGKTVGSSRFQATQCRQKRVILVREAPS--DA 2804  
 DB 160 CSTPYVCLQEQVRLQWQGDPPAR-----SVTFNSQKFEPTGVGH--LETLHMANWSQDH 211  
 QY 2805 GEVV---FSVRG--LTSKASLIVRERPAAI---IKPLEQOWVAPGDVELRELSEL--A 2853  
 DB 212 GRILCOLSVANHRAQSEIHLQVAPKGVKILLSP--SGRNILPGLVTLTCQVNSSYPA 270  
 QY 2854 GTPVHMLKRAIKRSQYDVCEGTWAMLVIRGASLKDAGEYTCBEVEASKT-----AS 2908  
 DB 271 VSSIKWLKD--GVRLQTKGV-----LHLPQAAMSAGVYTQCAENGVSGLSVPPIS 320  
 QY 2909 LHVEKANCETBELTNLQVEEGK-----TAVPTCKTEHPAAT---VTWRKGLLELRASG 2959  
 DB 321 LHI-----FMAEVQVSPAGPILLENQTVTLVNTPEAPSDLRYSWKYKNHVLLEDAH 371  
 QY 2960 KHQPSQSGTLRTISALEKADSDTYTCDIGQAQSAQLLVQGRRVHITLEDVDVQEG 3019  
 DB 372 SH-----TLRLHLAT--RADTGFCEV-----QNVHSGERSGPNVSV--- 406  
 QY 3020 SSATFCRISRPANYPVHWFELDKTP-----LHANELNE-----IDAQPGGY 3060  
 DB 407 -----VNNHPL-----TPVLTAFLQAGLVGILHCVSVSEPLATLVLSHGG- 449  
 QY 3061 HVLT-----LROLAKDSGTYIEAGDORASA----- 3087  
 DB 450 HILASTSGSDHSPRSGTSGPNSLRLEIRDEETDSGEYKCSATNSLGNATSTLDFHAN 509  
 QY 3088 ALRVTEKPSVFSRELTDAITEGEDTLVCETS---TCDIPMCWTQKDGKTLRGSARCQLS 3144  
 DB 510 AARLLTSPA-----AEVVEGQAVTLSCRSGLSPTPDARFSWYINGALL----- 552  
 QY 3145 HECHRAQLLTGATLQDSGRYKC-----EAGGACSSSIVRV--HARVRFQEAALKDLEV 3196

DB 553 HEGPGSSLLPLPAASSTDAGSVHCRARDGSHSAGSPSPAVLTVLPYPPROPTTTL--DUDA 611  
 QY 3197 LEGGA---ATLRC-VLSSVAAPVKWYGNVLRP-----GDKYSLRQEGAM 3238  
 DB 612 ACAGAGRRGILLCRVDSPPPARLQLLHLDVRVATSLPSGGGSCGCGSPMKVTKAENL 671  
 QY 3239 LELVVRNLRPDQSGRYSC-----SFGDQTSATL-----TVTAL-PAOFICKLRNKEATEGA 3289  
 DB 672 LRVEITHNLLBEEGLYLCEASNALGNASTATFNGQATVLAIAFSHTL-----QEGT 723  
 QY 3290 TATLRCELSKTAPEVWRKGSSET---LRDGRYCLRQDQAMCELOIRGLAMVDAEYSCV 3345  
 DB 724 EANTLCNVSREA-----AGSPANFSWFNG---VLMAQGPLETVTLLPVARTDAALYACR 775  
 QY 3346 CGEE-----RTSASL-----TIRPMAHFIGLRHOES 3373  
 DB 776 ILTEAQAQLSTPVLVSVLPDPKLSALLDMQGHMALFICTVDSRPLALLA-LFHGEH 834  
 QY 3374 IEGATATLRCELSKAAPVWRKGRSLRDGRHSLRODQAVCELOICGLAVADAGEYSC- 3432  
 DB 835 L-----LATSGLQVVP-----SHGRFOKAEANSLKLEVERELGLGDSGSVRCE 877  
 QY 3433 ---VCGEERTSATLTVKALPAKFTTEGLRNEEPAVEGATAMLWCELSKVAP-----VEMRKG 3485  
 DB 878 ATNVLGSSNTSLFFQVR---GAMVQVSPSPPELQGGQAVLSCQVHTGVEGTSYRW--- 930  
 QY 3486 ENLRDGRYILRQSGTRCELOICGLAMADAGEYLCVCGOERTSATLIRA-----LPA 3538  
 DB 931 ---YRDGQPL---QESTSATLFAAITLTQAGAYHCQ-AQAPGSATTSLAAPISLHVSYAP 984  
 QY 3539 RFIEDVRKNQEAEGATAVLOCELNSAAPVWRKGSSETLRDGRY---SLRQDG----- 3588  
 DB 985 RHVTLTLMDDTPGRGLGLLLCRVSDPPAQLR-----LLHGDRLVASTLQGVGSGESSP 1039  
 QY 3589 -----TKCELOIRGLAMADTGEYSC---VCGOERTSAMLTVRALPKFTFEGLEGRNEE 3636  
 DB 1040 RLHVAVAPNTLRLIEHGMLEDEGVYICEASNTLQASASADFDAQAVNVQVWPVGP---AT 1096  
 QY 3637 ATEGATAVLRCELSKMAPE---WVK-GHETLRDGRHSLRODQARCELOIRGLVAEDA 3691  
 DB 1097 VREGVLNLTCLVWTHPAQLTYTWYDGGQRL---DAHS-----IPLPNVTVRDA 1144  
 QY 3692 GEY:CMCKERTSAMLTVRAMPKFIIEGLRN-----EATEGDTATLWCELSKAAPVW 3745  
 DB 1145 TSYRCGVGPGRAPRLS-RPITLDLVYAPRNRLTYLLESHGGQLALVLTCTVDSRPPAQL 1203  
 QY 3746 RKGH-----ETLRDGRHSLRQDGSRCCELOIRGLAVVDAGYSCVC---G 3787  
 DB 1204 ALSHAGRLLASSTAASVNTLR-----LELRGPQPRDEGYSCARSPLG 1248  
 QY 3788 QERTSATLTVRALPARFIEDVKMQEAR--EGATAVLOC-ELSKAAPV--EWRKGSSETLRG 3842  
 DB 1249 QANTSLERLEG-----VRVILAPEAAVPEGAPITVTCADPAHAAPTLYTWYHNGRWL-- 1301  
 QY 3843 GDRYSLRQDQTRCELOIHGLSVADTGEYSCVC-----GQERTSATLTVPAQPVFREPL 3896  
 DB 1302 -----QEGPAASLSFLVATRAHAGAYSQAOAQGTRSSRPAALQVLVAPQDAVLSSP 1354  
 QY 3897 QSLQAEBSSTATLOCEL-SEPTA-----TVWMSKGLQLOANGRRERPRLOGCTAELV 3947  
 DB 1355 RDSRAR--SNAVIQCTVDSPPAELALSHDGKVLATSSGVHSLASGTGHVQVARNALRLQ 1412  
 QY 3948 LODLQREDTGEYTCCTC-----GSQATSATLTVTAAPVRFELQHQVDEGGTAHLCELS 4003  
 DB 1413 VQDVPAQD-DTYVYCTAQNLLGSIITIGRLQVEGA--RVVAE-PGLDVPEGAALNLSRLL 1468  
 QY 4004 RAGASVE-----WRKGSLLQFPFCAKYQMVQDAAEALLVRGVEQEDADGYTC---DT 4052  
 DB 1469 GGPFPVGNSTFAWFWNDRRLHASP-----VPTLAFTHVARAQAGWYHCLAEPLT 1517  
 QY 4053 GHTQSM-ASLSVRVPRPKFTRLOSLEOETGDTARLCCQLSDAESGAVVQWLKEGVELHA 4111



Db 1518 GAASAPVMLRLVLP-PKTPMVFVEGGLRGILDCRV-DSEPLA-----SUTLHL 1568  
QY 4112 GPKYEMRS--QGATRELLIH-----QLEAKDTGEVAC-----VTGGOKTA 4149  
Db 1569 GSRIVASSQPGCAPAEAPHIVLASPNALRVVDIEALRPSDQGEVICSASNVLSASTS 1625

## RESULT 11

MYPC HUMAN STANDARD; PRT; 1274 AA.  
ID QYPC HUMAN Q9UE27;  
AC Q14856; Q14856; Q9UE27;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Myosin-binding protein C, cardiac-type (Cardiac MyBP-C) (C-protein,  
DE cardiac muscle isoform).  
GN MYBPC3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Heart;  
RX MEDLINE=95262635; PubMed=7744002;  
RA Gautel M., Zuffardi O., Freiburg A., Labelt S.;  
RT "Phosphorylation switches specific for the cardiac isoform of myosin  
RT binding protein-C: a modulator of cardiac contraction?";  
RL EMBO J. 14:1952-1960(1995).  
RN [2]  
RP SEQUENCE FROM N.A. AND VARIANT CMH4 GLN-542.  
RX MEDLINE=97200835; PubMed=9048664;  
RA Carrier L., Bonne G., Bahrend E., Yu B., Richard P., Niel F.,  
RA Hainque B., Craud C., Gary F., Labelt S., Bouchour J.-S., Dubourg O.,  
RA Desnos M., Hagege A.A., Trent R.J., Komajda M., Fiszman M.,  
RA Schwartz K.;  
RT "Organization and sequence of human cardiac myosin binding protein c  
RT gene (MYBPC3) and identification of mutations predicted to produce  
RT truncated proteins in familial hypertrophic cardiomyopathy";  
RL Circ. Res. 80:427-434(1997).  
RN [3]  
RP SEQUENCE OF 640-694 FROM N.A.  
RX MEDLINE=96083593; PubMed=7493026;  
RA Bonne G., Carrier L., Bercovici J., Craud C., Richard P., Hainque B.,  
RA Gautel M., Labelt S., James M., Beckmann J.;  
RT "Cardiac myosin binding protein-C gene splice acceptor site mutation  
RT is associated with familial hypertrophic cardiomyopathy";  
RL Nat. Genet. 11:438-440(1995).  
RN [4]  
RP VARIANT CMH4 LYS-755.  
RX MEDLINE=98200174; PubMed=9541104;  
RA Yu B., French J.A., Carrier L., Jeremy R.W., McTaggart D.R.,  
RA Nicholson M.R., Hamby B., Sensarian C., Richmond D.R., Schwartz K.,  
RA Trent R.J.;  
RT "Molecular pathology of familial hypertrophic cardiomyopathy caused by  
RT mutations in the cardiac myosin binding protein C gene";  
RL J. Med. Genet. 35:205-210(1998).  
RN [5]  
RP VARIANT CMH4 HIS-654.  
RX MEDLINE=98200185; PubMed=9541115;  
RA Woolman-Smoock J.C., Mayosi B., Brink P., Corfield V.A.;  
RT "Identification of a new missense mutation in MYBPC-C associated with  
RT hypertrophic cardiomyopathy";  
RL J. Med. Genet. 35:253-254(1998).  
CC -!- FUNCTION: THICK FILAMENT-ASSOCIATED PROTEIN LOCATED IN THE  
CC CROSSBRIDGE REGION OF VERTEBRATE STRIATED MUSCLE. IN VITRO  
CC IT BINDS MHC, F-ACTIN AND NATIVE THIN FILAMENTS, AND MODIFIES THE  
CC ACTIVITY OF ACTIN-ACTIVED MYOSIN ATPASE. IT MAY MODULATE MUSCLE  
CC CONTRACTION OR MAY PLAY A MORE STRUCTURAL ROLE.  
CC -!- PTM: SUBSTRATE FOR PHOSPHORYLATION BY PKA AND PKC. REVERSIBLE  
CC PHOSPHORYLATION APPEARS TO MODULATE CONTRACTION (BY SIMILARITY).  
CC -!- DISEASE: Defects in MYBPC3 are the cause of familial hypertrophic  
CC cardiomyopathy type 4 (CMH4) [MIM:115197]; also known as FHC type

CC 4. CMH4 is an autosomal dominant disorder. It is a primary cardiac  
CC disease with risk of sudden unexpected death and is characterized  
CC by an unexplained hypertrophied non-dilated ventricle and by  
CC myocytic and myofibrillar disarray.  
CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. MyBP  
CC family.  
CC -!- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.  
CC -!- SIMILARITY: Contains 3 fibronectin type III domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X84075; CAA58882.1; -;  
CC EMBL; Y10129; CAA71216.1; -;  
CC EMBL; S80778; AAB35662.1; -;  
CC FIR; S55050; S55050.  
CC DR HSSP; P56276; ITLK.  
CC DR Genew; HGNC:7551; MYBPC3.  
CC DR MIM; 600958; -;  
CC DR MIM; 115197; -;  
CC DR GO; GO:0008307; P:structural constituent of muscle; TAS.  
CC DR GO; GO:0008936; P:muscle contraction; TAS.  
CC DR GO; GO:0008942; P:regulation of striated muscle contraction; TAS.  
CC DR InterPro; IPR008957; FN\_III-like.  
CC DR InterPro; IPR003961; FN\_III.  
CC DR InterPro; IPR007110; IG-like.  
CC DR InterPro; IPR003598; IG\_c2.  
CC DR Pfam; PF00041; f03; 3.  
CC DR Pfam; PF00047; ig1; 7.  
CC DR SMART; SM00060; FN3; 3.  
CC DR SMART; SM00408; IGc2; 1.  
CC DR PROSITE; PS50835; IG\_LIKE; 6.  
CC KW Immunoglobulin domain; Actin-binding; Cell adhesion; Muscle protein;  
KW Thick filament; Repeat; Phosphorylation; Disease mutation;  
KW Cardiomyopathy; Polymorphism.  
FT DOMAIN 102 152 PRO-RICH.  
FT DOMAIN 153 256 IG-LIKE C2-TYPE 1.  
FT DOMAIN 257 352 IG-LIKE C2-TYPE 2.  
FT DOMAIN 353 452 IG-LIKE C2-TYPE 3.  
FT DOMAIN 453 543 IG-LIKE C2-TYPE 4.  
FT DOMAIN 544 633 IG-LIKE C2-TYPE 5.  
FT DOMAIN 634 771 FIBRONECTIN TYPE-III 1.  
FT DOMAIN 772 869 FIBRONECTIN TYPE-III 2.  
FT DOMAIN 870 970 FIBRONECTIN TYPE-III 3.  
FT DOMAIN 971 1085 IG-LIKE C2-TYPE 6.  
FT DOMAIN 1086 1180 FIBRONECTIN TYPE-III 3.  
FT DOMAIN 1181 1274 IG-LIKE C2-TYPE 7.  
FT MOD\_RES 275 275 PHOSPHORYLATION (BY PKA AND PKC) (BY  
FT SIMILARITY).  
FT MOD\_RES 284 284 PHOSPHORYLATION (BY PKA AND PKC) (BY  
FT SIMILARITY).  
FT MOD\_RES 304 304 PHOSPHORYLATION (BY PKA AND PKC) (BY  
FT SIMILARITY).  
FT VARIANT 542 542 E -> Q (in CMH4).  
FT VARIANT 654 654 /FTIG=VAR 003917.  
FT VARIANT 755 755 R -> H (in CMH4; dbSNP:1800365).  
FT VARIANT 755 755 /FTIG=VAR 003918.  
FT VARIANT 755 755 N -> K (in CMH4).  
FT CONFLICT 536 536 /FTIG=VAR 003919.  
FT CONFLICT 536 536 R -> A (IN REF. 2).  
SQ SEQUENCE 1274 AA; 140833 MW; D31FOEA1572E5417 CRC64;  
Query Match 1.4%; Score 571; DB 1; Length 1274;  
Best Local Similarity 21.3%; Pred. No. 2.5e-12;  
Matches 336; Conservative 176; Mismatches 585; Indels 478; Gaps 63;  
QY 3804 FIEDVKNQFAREGATAVLCQELSKA-APVEMKSGSETLRGDRYSLRQDGTCEIQHGL 3862  
Db 12 FSKPKRSVEAAGSPAVFEATERAGVKVWRQGRGDSISANKYGLATEGTRHTLTREV 71

QY 3863 SVADTGBYSCVCGQERTSATLV-----RAPQVREPLQS 3898  
 Db 72 GPADQGYAVIAGSKYKFDLKVTEAKAPMLAPAPAEATGAGEAPAPA-AELGES 130  
 QY 3899 LOAEEGTATLQCELSPT-----A 3918  
 Db 131 APSPKGSSS---AALNGPTGAPDDPIGLFMRPQDGEVTVGGSITFSARVAGASLLKPP 187  
 QY 3919 TVVWSKG-----GLQOANGREPRLOGCTAEALVLODLOREDTGEVTCGCS---- 3965  
 Db 188 VVKFKKWDLKSVQCHLQHDSDYDRASKVLYFLHITDAQPTGSRVCEVSTKDXP 247  
 QY 3966 QATSATLTVTAAPVRFURELHQEVEDEGGTAHLCC--ELSRAG----- 4006  
 Db 248 ECSNFLTV-----HEAMGTGDLLSAFRRTSLAGGRRISDSHEDTGILDF 295  
 QY 4007 ASVEWRKGSLLQFPKAKYQMVQDGAALVVRGVEQEDAGDY---TCDTGTQSMASL-- 4061  
 Db 296 SSLKKRDSFRTPRDSKLEAPAEEDVWEIL---RQAPPSEYBRIAFQYGVTDLRGMLKR 351  
 QY 4062 --SVRVRPK---PKTRLQSLQ-ETGDIARLCCQLSDAESGAVQWOLKEGVELH-AGPK 4114  
 Db 352 LKGMREDEKSTAFQKLEPAYQVSKGHKIRLTVELADHD--AEVKLXNGQEIQMSGSK 409  
 QY 4115 YKRSQATRELLIHLEAKDTEYACVGTGQKTAASLRVTEPEVTVIRGLVDAEVTABE 4174  
 Db 410 YIFESIGAKRTLITISQCLADDAAYQCWVGGEKSTELFKPEPVLITRPLEDQLVWVGQ 469  
 QY 4175 DVFECSVRAGATGVQWQLQGLQSLQNEVTEVAV-RDGRHTRLRKLVTPEDAGTVSFH 4233  
 Db 470 RVEFECEVSEGAQ-VKWLKXGVELTREFTKFRFKXGQGRHLIINEAMLEDAQHYALC 528  
 QY 4234 LGNHASSAQITVRAPEVITILEPLQVQLSEGDQASFCQLSRASQGEAR--NALGG---V 4288  
 Db 529 TSGQALRELIVQEKLEVTQSIADLMVGAQDAQVEKCEV---SDENVRGWMLKNGKELV 585  
 QY 4289 PLOANEMNDITVEGTHLHLTHKVTLEDAGTVSFHVGTCSSEAQLKVTAKNTVVRGLEN 4348  
 Db 586 PDSRIKVSHT---GRVHKLTIDDTVPADAEVSYF----- 616  
 QY 4349 VEALGEGEALFECQLSQSOPEVAATHWLLDDEPVRTSENAEYVFFPENGRLHLLKLNLRPOD 4408  
 Db 617 --VPEG---FACNLS-----AKLHWE-----VKIDFVPRQES 643  
 QY 4409 SCRVTFLAGDMVTSAFITVVGWRLILEPLKNAAVRAGAQAQRTCTLSAEPVGERASWTI 4468  
 Db 644 PKIHLDCGRIPDTIVVAGNKLRLDVPI-----SGDPAP-TVIWQKAITQG----- 690  
 QY 4469 NGAAVOP-----DDSDWT-----VTADGSHQALLRSAPHPHAGEVT 4505  
 Db 691 NKAPARAPADPDTGSDSEWFDKLLCBTEGRVVRVETTKORSITVGEAKEDSGVYT 750  
 QY 4506 PACRDVARSALT---VLGLPDPPEAEVVAHSSHTVLSWAAPMSDGGGLCGVRVRE 4561  
 Db 751 VTKNPNVGEDQVNLTKVIDVPDPAAPAKTSNVEDSCTVQWEPFAYDGGQFILGYLER 810  
 QY 4562 KEGATGQWR---LCHELVPGCEVVDGLAPGETYRFRVAAGVPVGAEPVHLPTQVRL 4616  
 Db 811 KKKSYEMQNLNFDLIQELSHEARMEGV---VYEMRVYVNALGMSRPPASQ---- 862  
 QY 4617 AEPKPVPPQSPAPESQVAAAGEDVLSLELVAEAGEVIVHKGMERIQPGGRFVVVSQGR 4676  
 Db 863 --PFMPITGP-PSEPTHLAVEDVSDTTVSLK-----W-RPPEVVGAGG----- 900  
 QY 4677 QQMLVKGFAEAOQGEVHCGLAQSICPAAATFOVALSPASVDEAQPSPREAAQEGDL 4736  
 Db 901 -----LDGYSVE-----YC-----PEGCS----- 915  
 QY 4737 HLLWEALARKRMSREPTLDSISLPEDEGRSQRLPQEAEBEAVPDLSEGYSTADELARTG 4796  
 Db 916 ---WVA-----ALQGLTE----- 925

QY 4797 DADLSHTSSDDESAGTSLTYLTKKAGRPCTSLASKVKGAPAAP--SVKFPQOOQELPAA 4854  
 Db 926 -----HTS-----ILVKDLPTGARLLFRVRAHNMACGAPVITTEPVTVOEILQ- 969  
 QY 4855 VRPPIGDLSTKDLGDPMDKAAVKIQAFAKGYKVRKEMKQOEGPMFSGHTFGDTEAQVGDA 4914  
 Db 970 -RPRL-----QLFRHLRQ-----TIQKQVGE 990  
 QY 4915 LRLECVAASKADVARMLKQGVLTGRRHHHIDQGDGTCSLLIAGLDRADAGCY--TCQ 4972  
 Db 991 VNLIPFQCKPRPQVTTWKSCQPLA-GEVSIRNSPTDTI-LFIRAARRVHSGTYQVTVR 1048  
 QY 4973 VSNKFGQVTHSACVVVSGSESEAESESSGGELDDAFRAARLRLHRLFTKSPAENVSDBELF 5032  
 Db 1049 IENMEDKATLVQVVDKPPSPQ-----DLRVTDAWGLNVALEWKPKQPDVGNTELM 1098  
 QY 5033 ---LSADEGPAEPEEPADQTYREDEHFICIRFEALTEARQAVTRFOEMFATIGIVEI 5088  
 Db 1099 GYTVQKAD-----KKNWEW-----FTVLEHYRTHCVVPELI--LNGYYF 1137  
 QY 5089 KLVEQGPRRV--EMCISKE---TPAVVPPEPFLSLTSD--AAPVFLTELQNOEVQDGY 5141  
 Db 1138 RVFSQNMVGFSDRAATTKPEVFIPIRPGITYPEP-FNYKALDFSEAPSTQPLVNSVIAGY 1196  
 QY 5142 PVSFDCVVTGQPMFSVNFKDKLLEDDHYMINEDOGGHOLIITAVPADMGVYRCLA 5201  
 Db 1197 TAMLCCAVRGSFPKISFNKNGLDGEDARFM-FSQGVLTLEIRKPCPPDGGIYVCRA 1255  
 QY 5202 ENSMGVSSTKAELRV 5216  
 Db 1256 TNLQGEARCECELEV 1270

RESULT 12  
 SN MOUSE  
 ID SN MOUSE STANDARD; PRT: 1694 AA.  
 AC Q62230; O55216; Q62228; Q62229;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Sialoadhesin precursor (Sialic acid binding Ig-like lectin-1) (Siglec-1) (Sheep erythrocyte receptor) (SER).  
 GN SN OR SA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euhelzia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=C57BL/6; TISSUE=Macrophage;  
 RX MEDLINE=95009950; PubMed=7925291;  
 RA Crocker P.R.; Mucklow S.; Boukson V.; McWilliam A.; Willis A.C., Gordon S.; Milon G.; Keim S.; Bradford P.;  
 RT "Sialoadhesin, a macrophage sialic acid binding receptor for haemopoietic cells with 17 immunoglobulin-like domains".  
 RL EMBO J. 13:4490-4503 (1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95051930; PubMed=9383289;  
 RA Mucklow S.; Gordon S.; Crocker P.R.;  
 RT "Characterization of the mouse sialoadhesin gene, Sn.";  
 RL Mamm. Genome 8:934-937 (1997).  
 RN [3]  
 RP CHARACTERIZATION.  
 RC STRAIN=C57BL/6; TISSUE=Spleen;  
 RX MEDLINE=91266893; PubMed=2050106;  
 RA Crocker P.R.;  
 RT "Purification and properties of sialoadhesin, a sialic acid-binding receptor of murine tissue macrophages".  
 RL EMBO J. 10:1661-1669 (1991).  
 RN [4]  
 RP SIALIC ACID BINDING.  
 RX MEDLINE=95179521; PubMed=7533044;

RA Kelm S., Pelz A., Schauer R., Filbin M.T., Tang S., de Bellard M.E.,  
 RA Schnaar R.L., Mahoney J.A., Hartnell A., Bradfield P., Crocker P.R.;  
 RT "Sialoadhesin, myelin-associated glycoprotein and CD22 define a new  
 RT family of sialic acid-dependent adhesion molecules of the  
 RL immunoglobulin superfamily.";   
 RL Curr. Biol. 4:965-972(1994).  
 RN BINDING TO SPN.  
 RP MEDLINE=21136329; PubMed=11238599;  
 RA van den Berg T.K., Nath D., Ziltener H.J., Vestweber D., Fukuda M.,  
 RA "CD43 functions as a T cell counterreceptor for the macrophage  
 RT adhesion receptor sialoadhesin (Siglec-1).";  
 RL J. Immunol. 166:3637-3640(2001).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 20-138.  
 RX MEDLINE=98325385; PubMed=9660955;  
 RA May A.P., Robinson R.C., Vinson M., Crocker P.R., Jones E.Y.;  
 RT "Crystal structure of the N-terminal domain of sialoadhesin in  
 RT complex with 3' sialyllactose at 1.85 A resolution.";   
 RL Mol. Cell 1:719-728(1998).  
 RN [7]  
 RP STRUCTURE OF 20-138 BY NMR, AND MUTAGENESIS OF TRP-21 AND ARG-116.  
 RX MEDLINE=99321481; PubMed=10393093;  
 RA Crocker P.R., Vinson M., Kelm S., Drickamer K.;  
 RT "Molecular analysis of sialoside binding to sialoadhesin by NMR and  
 RT site-directed mutagenesis";   
 RL Biochem. J. 341:355-361(1999).  
 CC -!- FUNCTION: Macrophage-restricted adhesion molecule that mediates  
 CC steric-acid dependent binding to lymphocytes, including  
 CC granulocytes, monocytes, natural killer cells, B-cells and CD8 T-  
 CC cells (By similarity). Preferentially binds to alpha2,3-linked  
 CC sialic acid. Binds to SPN/CD43 on T-cells. May play a role in  
 CC hemopoiesis.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (isoform 1) and  
 CC soluble (isoforms 2 and 3).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=1;  
 CC IsoId=Q62230-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q62230-2; Sequence=VSP\_002573, VSP\_002574;  
 CC Name=3;  
 CC IsoId=Q62230-3; Sequence=VSP\_002575, VSP\_002576;  
 CC -!- TISSUE SPECIFICITY: Expressed by macrophages in various tissues.  
 CC Highest expression in spleen and lymph node with lower amounts in  
 CC lung, liver, bone marrow, heart and skin. No expression in thymus,  
 CC kidney, brain or small intestine.  
 CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC  
 CC (sialic acid binding Ig-like lectin) family.  
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC -!- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z36293; CA855290.1; -;  
 DR EMBL; Z36233; CA855268.1; -;  
 DR EMBL; Z36234; CA855269.1; -;  
 DR EMBL; U92842; AAB95641.1; -;  
 DR EMBL; U92833; AAB95641.1; JOINED.  
 DR EMBL; U92834; AAB95641.1; JOINED.  
 DR EMBL; U92836; AAB95641.1; JOINED.  
 DR EMBL; U92837; AAB95641.1; JOINED.  
 DR EMBL; U92838; AAB95641.1; JOINED.  
 DR EMBL; U92839; AAB95641.1; JOINED.  
 DR EMBL; U92840; AAB95641.1; JOINED.  
 DR EMBL; U92841; AAB95641.1; JOINED.

DR PIR; S50065; S50065.  
 DR PDB; 1QFO; 16-APR-99.  
 DR PDB; 1QFP; 16-APR-99.  
 DR MGD; MGI:99668; Sn.  
 DR GO; GO:0016021; C:integral to membrane; ISS.  
 DR GO; GO:0005529; F:sugar binding; ISS.  
 DR GO; GO:0018337; P:cell-cell adhesion; ISS.  
 DR GO; GO:0007160; P:cell-matrix adhesion; ISS.  
 DR GO; GO:0006954; P:inflammatory response; ISS.  
 DR InterPro; IPR007110; IG\_c2.  
 DR InterPro; IPR003598; IG\_c2.  
 DR Pfam; PF00047; ig\_16.  
 DR SMART; SM00408; IGc2; 5.  
 DR PROSITE; PS00835; IG\_LIKE; 14.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 KW Cell adhesion; Lectin; Transmembrane; Signal; Glycoprotein;  
 KW Immunoglobulin domain; Repeat; Alternative splicing; 3D-structure.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1694 SIALOADHESIN.  
 FT DOMAIN 20 1638 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1639 1659 POTENTIAL.  
 FT DOMAIN 1660 1694 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 20 136 IG-LIKE V-TYPE.  
 FT DOMAIN 153 235 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 239 321 IG-LIKE C2-TYPE 2.  
 FT DOMAIN 326 406 IG-LIKE C2-TYPE 3.  
 FT DOMAIN 416 508 IG-LIKE C2-TYPE 4.  
 FT DOMAIN 509 594 IG-LIKE C2-TYPE 5.  
 FT DOMAIN 602 701 IG-LIKE C2-TYPE 6.  
 FT DOMAIN 704 781 IG-LIKE C2-TYPE 7.  
 FT DOMAIN 795 890 IG-LIKE C2-TYPE 8.  
 FT DOMAIN 894 972 IG-LIKE C2-TYPE 9.  
 FT DOMAIN 979 1078 IG-LIKE C2-TYPE 10.  
 FT DOMAIN 1080 1160 IG-LIKE C2-TYPE 11.  
 FT DOMAIN 1171 1263 IG-LIKE C2-TYPE 12.  
 FT DOMAIN 1244 1336 IG-LIKE C2-TYPE 13.  
 FT DOMAIN 1341 1438 IG-LIKE C2-TYPE 14.  
 FT DOMAIN 1441 1519 IG-LIKE C2-TYPE 15.  
 FT DOMAIN 1533 1626 IG-LIKE C2-TYPE 16.  
 FT SITE 827 829 CELL ATTACHMENT SITE (POTENTIAL).  
 FT DISULFID 41 98 BY SIMILARITY.  
 FT DISULFID 160 218 BY SIMILARITY.  
 FT DISULFID 263 306 BY SIMILARITY.  
 FT DISULFID 347 391 BY SIMILARITY.  
 FT DISULFID 434 492 BY SIMILARITY.  
 FT DISULFID 532 576 BY SIMILARITY.  
 FT DISULFID 625 685 BY SIMILARITY.  
 FT DISULFID 725 770 BY SIMILARITY.  
 FT DISULFID 813 872 BY SIMILARITY.  
 FT DISULFID 911 955 BY SIMILARITY.  
 FT DISULFID 1000 1062 BY SIMILARITY.  
 FT DISULFID 1102 1144 BY SIMILARITY.  
 FT DISULFID 1188 1236 BY SIMILARITY.  
 FT DISULFID 1276 1319 BY SIMILARITY.  
 FT DISULFID 1362 1421 BY SIMILARITY.  
 FT DISULFID 1462 1508 BY SIMILARITY.  
 FT DISULFID 1551 1610 BY SIMILARITY.  
 FT CARBOHYD 159 159 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 266 266 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 500 500 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 583 583 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 693 693 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 722 722 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 737 737 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 882 882 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1099 1099 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1246 1246 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1459 1459 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1473 1473 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT VARSPLIC 326 340 MAEVKMPAPVLEN -> SESWMLRGVPVSGKH (in isoform 2). /FTID=VSP\_002573.  
 FT VARSPLIC 341 1694 Missing (in isoform 2). /FTID=VSP\_002574.  
 FT VARSPLIC 1528 1598 LVASNLHDAPTKPHIRVAPPNALRDVIE -> CEVEPIS  
 FT ALCLSLHGTGPYAFSSAQSGKGLRGLASSLACWFF  
 FT VSMGLYPALKWRLLPFWDEYR (in isoform 3).  
 FT /FTID=VSP\_002575.  
 FT VARSPLIC 1599 1694 Missing (in isoform 3).  
 FT /FTID=VSP\_002576.  
 FT MUTAGEN 21 21 W->O: LOSS OF SIALIC ACID BINDING.  
 FT MUTAGEN 116 116 R->A: LOSS OF SIALIC ACID BINDING.  
 FT MUTAGEN 116 116 R->L: 10-FOLD LOSS IN AFFINITY TO SIALIC

Query Match 1.3%; Score 532; DB 1; Length 1694;  
 Best Local Similarity 20.7%; Pred. No. 7.5e-11;  
 Matches 416; Conservative 240; Mismatches 691; Indels 660; Gaps 100;

QY 790 SVQGLARFLKDMAGSC--VDVAGGPAQPECETSEAHVHVWYKD-----GMELHSGGR 843  
 DB 28 NVQGL-----SGSCLLIPCIPSPADVPVNG-----ITAIWYDYSGRQVVIHSGDP 76  
 QY 844 FLOEDVGRHR-----LVAATVTRQDEGTYSRCVBGSDVDRLRVSE--- 885  
 DB 77 KL--VDKRFGRALMGNDHKVNCNLLDLKPEDSGTY-----NFRFEISDSNR 124  
 QY 886 -----PKVYFAKQLARRKLOAEAGASATLSCEVAQAQTEVTWY--KDGK 928  
 DB 125 WLVKGTVTVTDPSPPTIPIEL-REGERNFNCTPYLC-LOEKQVSLQWRQDPT 182  
 QY 929 KLSSSKVCMEATCTRLRVVQQA-----GQADAGEYSCAAGQRLSFHLDVKE-PKVY 981  
 DB 183 HSVTSSQSLPTSGYHQTTLHMSLWQDHGRITLQFSLGAHSSRKEVVYLQVHPAPKV 242  
 QY 982 FAKQVHAHVEQAAGANATLSCEVAO---AQAEWYKDGKGLSSSLKVHVEAKGRRR 1038  
 DB 243 --EILLSSGNILPGDPVTLTKVNSSYPAVSAVOWARDGVNL--GVTHGV----- 290  
 QY 1039 LVQOAGKTADGYSCEA---RGQVRS--FRLHITEPKMFAKEQSVHNEVQAEAGASAM 1093  
 DB 291 LRLFAANWDSGATCOATNDMSLVSSPLSLHV-----FMAEVKMNAPGVLENEVT 344  
 QY 1094 LSCEV---AQAEVTVWYKDGKGLSSSKVGMVKGCTRLVLPQAKADAGEYSCA-- 1148  
 DB 345 LLCSTPKAEPOELRYSWKHILLLEDAH-----ASTLHLPVAVTRADTGFYFCEVQN 395  
 QY 1149 --GGQVRSFHLHITEPKGVFAKEQSVHNE-----VOAEAGTAMLSCE-VAQPQTEVTWYK 1201  
 DB 396 AQGERS-----PLSVVRYPLPTDITFLETOAGLVGLHCSVSEPLATVLSH 448  
 QY 1202 DGKLSSSS-----KYRMEVKGCTRLRVVQVQKADAGEYSCA-----GGQVRSFQLH 1250  
 DB 449 GGLTLASNGENDNPRFR--SSAPNSLRLEIRDLQPADSGEYTCVAVNSLGNSTSLDFY 508  
 QY 1251 ITEPKAVFAKQLVHNEVRIEAGASATLSCEVAQA---QTEVTWYKDGKL---SSSSKV 1304  
 DB 509 ANVARL-----LNPAAVEVQGAUTLSRSGLSPPADTRFRSWYINGALLLEGSSSS-- 560  
 QY 1305 RIEAAGCMQLVVOQAQADAGEYTC--EAG-----EAG----- 1333  
 DB 561 -----LLPAASTSDAGSYCYRTQAGPNTSGPSLPTVLTVPYPRKPTFTARLDL 610  
 QY 1334 -----QOR--LSFHLVDSEPKAVFAKEQLAHRKVQAEAGATLS-----CEVAQAQTE 1380  
 DB 611 DTSGVGGRGILLCHVDSDPP-----AQLRLHK-----GHVVA:SLPSRCGSCSQRTK 660  
 QY 1381 VTWYKDGKGLSSSKVAMEAVGCTRLRVVQQAQOATDGEYSCA-----GGQRLSFSLDVA 1436  
 DB 661 VS-----RTSNSLHVE-----IQPVLDEBGEYVLCASNTLGNSSAAASFNAK 703

QY 1437 BPKVFAKEQPVHREVQQAAGASTTISC-----EVAQAQTEVWYKDGKGLSSSKVME 1492  
 DB 704 ATVLVITSNTL-----REGTEANLTCNGQEVAVSPANFSWFRNGV-----LMT 748  
 QY 1493 VGCTRLRVVQQAQADAGEYSC-----EAGSORLSFHLHVAEP---KAVFAKEQ-ASREV 1544  
 DB 749 QGSLETVRQLLARTDAAVACRLLTEDQAQ-----LSAPVVLVLYADPPKLSALL 801  
 QY 1545 QAEAGTSATLSCEV-AQAQTEVTWYKDGKLS-----SSSKVMEAVGCTRLRVVQ 1595  
 DB 802 DVQGRMAVFICTVDVSYPLAHLISFRGDHLLATNLEPQPSHGRIQAKATANSIQLEVRE 861  
 QY 1596 AQQADAGEYSCA-----GDQRLSFHLHVAEPKVFAKEQPAHREVQAEAGASATLSCEVA 1651  
 DB 862 LGLVDGNGVHCENATNLSANSLSLFFQVGAWVFT-----ITELREGQAVVLSQV 914  
 QY 1652 QAEQTEVT---WYKDGKGL---SSSKVMEAVGCTRLRVVQQAQADAGEYSCAAGQRLS 1706  
 DB 915 TGVSEGTYSYWDGRPLQESTSTLRIAAISLRQ-----AGAVHCQA----- 957  
 QY 1707 FRLHVALEPQISERPCRRREPLVKEHEDIILTATLATPSAATVTWLKDGVEIRSRHE 1766  
 DB 958 -----QAPDTAI-ASLAAPVSLHVS-----TERHV 982  
 QY 1767 TASQGDTHLTVHGAQVLDLSAISYCRVGAEGQDPFVQVEEVAAKFCLLPEVCGELGTV 1826  
 DB 983 TLS-----ALLSTDPERLGHVCSVQ-----SDPPAQLQFHRN--RL---VASTLQAD 1037  
 QY 1827 TLACELSPACAEVWRCNGTQPRVGRFQVMAEGPVRSITVL--GLR-----AEDAGE 1877  
 DB 1028 ELA-----GSNPL-----HVTLPNELQLQTHFPELEDDGT 1059  
 QY 1878 YVCESRDHTSAQLTVSVRVVWFKMSGLSTVASEGGEATFCQV---SPSDVAVWFRD 1934  
 DB 1060 YTCESANTLGQASAAADFDAQAVRVVWPNATVQEGQVNLTCVWSTHODLSYTYWK 1119  
 QY 1935 GALLQSEKFAISQSGASHSLTISDLVLEDAQ---QITVEAEGASSAALRVREAPVLFK 1992  
 DB 1120 G-----QQLGARSITLPSVKVLDATSVRCVGLPG-----HAPHLSR- 1157  
 QY 1993 KLEPQTEBERSVTLEBELTFPPELWTRN--ATALAPKXVBIHAEGARHLRVLHVNG 2050  
 DB 1158 -----PVTLDV-----LHAPNRLUTYL-----LETQGRQLALVL--- 1187  
 QY 2051 FADRGFFCETFDKTOAKLTVMQRVRLVGLQAVEAREGQTA--TMEVQLSHADVDS 2108  
 DB 1188 -----C-TVDSRPPAQLT-----LSHGDQVVASSTEASVNTLRUEL----- 1223  
 QY 2109 WTRDGLRFQCGPTCHLAVRGPMTLTLGLRPEDSGMLVFKAEQ-----VHTSARLVVTEL 2164  
 DB 1224 -----QDP-----RPSNEGLYSCSAHSPGLKANTISLELLEGV 1256  
 QY 2165 PVSFRLQDVVTTEKEKVTLECE--LSRPNVDVRLKDGVELBAGTKMAIAQAQACRL 2222  
 DB 1257 RVQMN-PSGSV--PEGEPTVTCEDPAALSALYAFWFGHWLQ-----EGPASSL 1304  
 QY 2223 TIYRCEFAQGVVYCDADHAQSSASVQVGRYTLIYRRVLAEDAGEIQFVAENAEARQ 2282  
 DB 1305 QFLVTRAHAGAYFCQVHDTQGTSSR--PASLQILY-----APRDAVLSSFRDSRTR 1355  
 QY 2283 LRV-----KELPVTILVRPLRDKIAMEKHR-----GVLEQVRSASAQVWFKGSGOEL 2329  
 DB 1356 LMWVTOCTVDSEPPAEVMSLHNGKVLAAHERHSASGIGHIQVARNAL----- 1404  
 QY 2330 QPGPKVELVSDGLYKLIISDVHAEDEDTYTCDA---GDVKTSQAQFFVEEQSITIVRGL 2385  
 DB 1405 -----RLQVQDVTILGDGNTYVCTAQHTLGLSISTTQRLITE-----TDIR-- 1443  
 QY 2386 QDVTWEPAPAWFCESTIPSVRPKMLLGKTVLQAGN----- 2424  
 DB 1444 --VTAEPGLDW-----PEGTALNLSCLLPGGSGPTGNSFTFWNRHRLHSA 1489  
 QY 2425 -----VGLEQSGTVHRLMLRITCTMTGTPVHFTVGKSSARLVVSDIPVVLTRPLE 2476

Db 1490 VPTLSFVPTVQAAGLHCRADLTGATTSAPVWMLRVLPKPTLIVF-----VE 1540  
 Qy 2477 PXTGRELSVVLSCDFRPAKAVQWYK-----DTPLSPSKFKXNSLEGQWAEAL 2525  
 Db 1541 PQGGHQ---GILDCRVDSPLAILTLHRGSLVASNLQHDAPTKP--HIRVTAPPNALRV 1595  
 Qy 2526 RILRLMPADAGVRCQA---GSAHSS 2548  
 Db 1596 DIEELGFSNQGEYVCTASNLUGSASAS 1622

RESULT 13  
 MYPF CHICK  
 ID MYPF CHICK STANDARD; PRT; 1131 AA.  
 AC P16419; Q90606; FT DOMAIN 47 148 IG-LIKE C2-TYPE 1.  
 DT 01-AUG-1990 (Rel. 15, Created) FT DOMAIN 248 337 IG-LIKE C2-TYPE 2.  
 DT 10-OCT-1996 (Rel. 34, Last sequence update) FT DOMAIN 338 428 IG-LIKE C2-TYPE 3.  
 DT 10-OCT-2003 (Rel. 42, Last annotation update) FT DOMAIN 429 529 IG-LIKE C2-TYPE 4.  
 DE MYosin-binding protein C, fast-type (Fast MyBP-C) (C-protein, skeletal muscle fast-isoform). FT DOMAIN 530 629 IG-LIKE C2-TYPE 5.  
 DE FT DOMAIN 630 727 FIBRONECTIN TYPE-III 1.  
 DE FT DOMAIN 728 828 FIBRONECTIN TYPE-III 2.  
 DE FT DOMAIN 829 922 IG-LIKE C2-TYPE 6.  
 DE FT DOMAIN 923 1037 FIBRONECTIN TYPE-III 3.  
 DE FT DOMAIN 1038 1131 IG-LIKE C2-TYPE 7.  
 DE FT SEQUENCE 1131 AA; 126811 MW; 73262DC5A939D6AC CRC64;  
 CN MYBPC2  
 OS Gallus gallus (Chicken).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-24; 182-198 AND 1030-1044.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=94043444; PubMed=8227129;  
 RA Okagaki T., Weber F.E., Fischman D.A., Vaughan K.T., Mikawa T., Reinach F.C.;  
 RA "The major myosin-binding domain of skeletal muscle MyBP-C (C protein) resides in the COOH-terminal, immunoglobulin C2 motif.";  
 RT J. Cell Biol. 123:619-626(1993).  
 RN [2]  
 RP SEQUENCE OF:139-1131 FROM N.A.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=90192766; PubMed=2315308;  
 RA Einheber S., Fischman D.A.;  
 RA "Isolation and characterization of a cDNA clone encoding avian skeletal muscle C-protein; an intracellular member of the immunoglobulin superfamily.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 87:2157-2161(1990).  
 RL CC -!- FUNCTION: THICK FILAMENT-ASSOCIATED PROTEIN LOCATED IN THE CROSSBRIDGE REGION OF VERTEBRATE STRIATED MUSCLE A BANDS. IN VITRO IT BINDS MHC, F-ACTIN AND NATIVE THIN FILAMENTS, AND MODIFIES THE ACTIVITY OF ACTIN-ACTIVED MYOSIN ATPASE. IT MAY MODULATE MUSCLE CONTRACTION OR MAY PLAY A MORE STRUCTURAL ROLE.  
 CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. MyBP family.  
 CC -!- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.  
 CC -!- SIMILARITY: Contains 3 fibronectin type III domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U00922; AAC59644.1; -  
 CC EMBL; U00923; AAC04307.1; -  
 CC EMBL; M31209; AAA49068.1; -  
 CC PIR; A48668; A35089.  
 CC RSP; P56276; 1TLK.  
 CC InterPro; IPR008957; FN III-like.  
 CC InterPro; IPR003961; FN\_III\_subd.  
 CC InterPro; IPR003962; FN\_III\_subd.  
 CC InterPro; IPR007110; Ig-like.  
 CC InterPro; IPR003598; Ig\_c2.

DR Pfam; PF00041; fn3; 3.  
 DR Pfam; PF00047; ig; 7.  
 DR PRINTS; PRO0014; FNTYPEIII.  
 DR SMART; SM00060; FN3; 3.  
 DR SMART; SM00408; IGc2; 1.  
 DR PROSITE; PS00835; IG LIKE; 5.  
 KW Immunoglobulin domain; Actin-binding; Cell adhesion; Muscle protein;  
 KW Thick filament; Repeat.  
 FT INIT MET 0  
 FT DOMAIN 47 148 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 248 337 IG-LIKE C2-TYPE 2.  
 FT DOMAIN 338 428 IG-LIKE C2-TYPE 3.  
 FT DOMAIN 429 529 IG-LIKE C2-TYPE 4.  
 FT DOMAIN 530 629 IG-LIKE C2-TYPE 5.  
 FT DOMAIN 630 727 FIBRONECTIN TYPE-III 1.  
 FT DOMAIN 728 828 FIBRONECTIN TYPE-III 2.  
 FT DOMAIN 829 922 IG-LIKE C2-TYPE 6.  
 FT DOMAIN 923 1037 FIBRONECTIN TYPE-III 3.  
 FT DOMAIN 1038 1131 IG-LIKE C2-TYPE 7.  
 FT SEQUENCE 1131 AA; 126811 MW; 73262DC5A939D6AC CRC64;  
 SQ  
 Query Match 1.3%; Score 528.5; DB 1; Length 1131;  
 Best Local Similarity 21.0%; Pred. No. 6e-11;  
 Matches 306; Conservative 172; Mismatches 523; Indels 453; Gaps 53;  
 QY 3817 ATAVLQCELSKAAPVWKR-----GSETLRGDRYSLRQDGTTC-ELQIHGSLVADTGE 3869  
 Db 73 AGAALPC-----APAVKFKGKMAELGDKSARCLRHSDVDDDKVHTFELTITKVMGDRGD 128  
 QY 3870 YSC-VGQER---TSATLTVRAPQVFPFQSL-QAEEGSTATLQCELSSEPTATVWNSK 3924  
 Db 129 YFCEVTAKEQKSCSFSDIIEAPRSEGNVLAQFRTGEGKDDT-----AGEUDFGLLKK 184  
 QY 3925 GGLQANGRRPR-----LQCTABLVLQDLQREDTGYTCTCGSQATSATLT 3973  
 Db 185 REVQVEKKKKDEDDQPPPEIWEELKGVTKKSEVERI---AFQGITDLRGLMKRLKK 240  
 QY 3974 VTAAPVR---FLRELQ-HQEVDEGGTAHLCCELSRAGASVEWVKSLQIFPCAQYQWQD 4029  
 Db 241 VHVPEKKEAFIRKLDPAQYQVQDKGNKILVWELSDPLPLKWKYKNGQLLKPTKTVFENV 300  
 QY 4030 GAAAEELVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFKTRLQSLQEQETGDIARLCC 4089  
 Db 301 GLKRLITHKSLADDAAYECVNDKCFTEVFVKEPPVTVVRGLEDQVVGVD--RVVL 358  
 QY 4090 QLSDAESGAVQWMLKEGVELHAGP--KYEMRSQGTARELLIHQLEAKDTGEVACTVGGQK 4147  
 Db 359 EAEVSEGAQVWMLKDGVDVTRDDAFKRFKDKGKKHFLIINEARLSDSAHYKIMTNGGE 418  
 QY 4148 TAASLRVTEPEVTIVRGLVDAEVTADEVDFESCEVSRAAGATGVQWCLQGLPLQSNVETEV 4207  
 Db 419 SEAELSVEEKQLEVLQDMADLTVKASQAVKCEVSDSEKVTG-RWFRNGVEVKPKRIHI 477  
 QY 4208 AVRDGRIRHLRLKGVTPEDAGTVSPHLGNHASSAQLTVRAPEVTILEPLQDVQLSEGQDA 4267  
 Db 478 S-HNGRFHKLVIDDVRPEDEGDYTFIPGYALSLSAKLNFLKIV-----521  
 QY 4268 SFQCLSRASGQEAQWALGVPLQANEMNDIIVEQGLLEHLLTHKVLLEDACTVSHVGT 4327  
 Db 522 -----EYVPKQ-----EPPKHL-----D 535  
 QY 4328 CSSEALQKVTAKNTVVRGLENVEALLEGALFECQLSQPEVAHAHTWLLDDEPVRTSENAE 4387  
 Db 536 CSGKA-----AENTIV-----VVGKNKRLVDVIGSEPAFTVTKSGDQLFTATE--- 580  
 QY 4388 VVFFENGLRHLLLLKRLNRQDSCRVTFLAGDMVTSAFITVRGWRLEILEPLKNAAVRAGA 4447  
 Db 581 -----GRVHI-----DS-----QADL--SSFVIESAERSD-----603  
 QY 4448 QARETCTLSEAVPVGEASWYINGAAVQPDSDSWTVTADGSHQALLLRGAQPHHAGEVTFA 4507  
 Db 604 EGRYCITVTN--PVGEDS-----619



QY 4011 WRK-----GSLQLPCCAKYQVQDGAALALLVROVQEDADGDTCDTGHQTQSWA 4059  
 Db 226 YEKIAFOYGITDLGMLKRLKAKAVEVKKSA-----257  
 QY 4060 SLSVRPRPKFKRLQSLQ-ETGDIALRCQLSDAESGAVQWLKGVVHAGPKYEMR 4118  
 Db 258 -----FTKLDPAVQVDRGNKIKLWVEISDPD--LTLKFKNGQEIKESSKYVPE 305  
 QY 4119 SQGATRELLIHQLEAKDTGAYCVTGQKNTAASLRVTEPEVTVIRGLVDAEVTADEDF 4178  
 Db 306 NVGKKRILITINKCTLADAAAYEAVKDEKCFTELKVEPPVLIPTPLEDQOVFVGRVEM 365  
 QY 4179 SCEVSRAGATGVQVCLQGLPLQSNVEVAVR--DGIHTLRLKGVTPEDAGTVFHLG 4235  
 Db 366 AVEVSEBGAQ-VMMKDGVELTREDSPKARYRFKDKGRHILFSDVVQEDRGYQVITN 424  
 QY 4236 NHASSAQLTVRAPEVITILEQLDQVQLSEGQDQSFQCRLS--RASGQEARWALGGVPLQAN 4293  
 Db 425 GGOCEAEELIVEEKQLEVLQDIADLTVKASEQAVFKCVSDEKVTG---KWKNGVEVRPS 481  
 QY 4294 EMNDITVEQ-GTLHLTLHLKLTLEDAGTVSF-----HVGTCSE--AQLKVTAKNTVVR 4344  
 Db 482 KR--ITISHVGRFKHLVIDVRPEDEGDYTFVPDGYALGSLSAKINFLKIV-----531  
 QY 4345 GLENVEALEGGEALFEQQLSQPEVAATHLLDDEPVRTSENAEVFFPENGRLHLLKLN 4404  
 Db 532 --EYVPKQEPKIHLDGS-----GXTSENAIVVAGNKLRL--LDVSI 569  
 QY 4405 RPQDSRCVTFIAGDMVTSAFITVGRWLEILEPLKNAAVRAGAQAQRTCTLSAEPVGEA 4464  
 Db 570 TGEPPVATWIKGDEV--FTTTEG-----RTIEKRYDC-----601  
 QY 4465 SWYINGAAVQDDSDWTVTADGSHQALLRSQAQPHAGEVTFACRDVAVASARLTVLGLPD 4524  
 Db 602 SSFVIEAQRDEGRYTI-----KVTNPVGEDVAFIFQVVDVDP 641  
 QY 4525 PPDAEVVAHSHSTVTLISWAAAPMDGGGLGGRYVVEKGGATQW-RLCHELVGPPECVV 4583  
 Db 642 PPEAVRITSVGEDNAILWVEPPMDGKPKVTGYLVERKKGQSWMKLNFVETETTTYES 701  
 QY 4584 DGLAPGETYRRAVAVGPVAGEFVHLPTQTVLAEPPKPPVPPQSPAPESQVAGEDVSL 4643  
 Db 702 TKMIEGLIYERFVAVNAIGVSQP-----SNTRPFPMIATTS-----EPHL 744  
 QY 4644 ELEVVAAEAGEVIMHKGMERIOFGG-----RFEVWSQGRQOMLVIKGFTAEDQGEYHCLGAQ 4699  
 Db 745 IVEDVTDITTLKWRPNRIGAGGIDGVLVEYCLEGSEEWV-----PANTEPVERCGFTV 799  
 QY 4700 GSICPAATFOVALSPASVDEAPQPSLPPAEQAQGDHLHLEALARKMRSEPTLDSIS 4759  
 Db 800 XNL-PTGARILFRVGVNIGARTFPA-----TLAQPTVIREIA 836  
 QY 4760 ELPEDGRSQRLPQAEAEVAPDLSEGYSTADELARTGDALSHTSDDSESRAGTFSLVTY 4819  
 Db 837 EPPK-----IRLPHLRQ-----TY 851  
 QY 4820 LKKA-----GRGTSPLASKVGAAPSVKPOQOQBPPLAARPPPLGLDUST-----4864  
 Db 852 IRKVGQELNLVVPFOCKPRPQVWTKGAPLDTS-----RVHVRTSDPTVFFV 900  
 QY 4865 -----KDLGDSMDKAAVKIQAFKGYKVRKEMQOQEGPMFSHTFGDTEAQVGDALALEC 4919  
 Db 901 ROARSDSGEVEL-----SVQIENMKDTATIRVVEKAGPPIN-----VMVKE 944  
 QY 4920 VVASKADYRARLWKDGVLTDRHHHIDQLGDGTCSLLIAGLDRADAGCYTCQVSNKFGQ 4979  
 Db 945 VWGTNALVWQAPKD-----DGNSEIMGYFVQKADK--KTMENFNVYER 986  
 QY 4980 VTHSACVVVSGSEASESSSGGELDDAFRAARLRLFRKTSAPAEVSDLEFLSADEGP 5039  
 Db 987 NRHTSCTV-----SLLIVGNEYFF-----1005

QY 5040 APEPSPADWQTYREDEHFICIRFEALTEARQAVTRFQEMFATLIGIGVIEKLVEQGPRRVE 5099  
 Db 1006 -----RVYTEN---IC-----GLSDSPGVSKNTARILKT---GITFKFEYKEHDFR 1046  
 QY 5100 MCISKETPAPVVPPEPLSLTSDAAPVFLTELQNEVQDGPVPSFDCVVTGQPMPSVRW 5159  
 Db 1047 M-----APKFLTPLIDRVVAVAGYSAALNCAVRGHPKPKVVM 1082  
 QY 5160 FKDGKLLBEDDHMYNEDQOGHQLIITAVVPADGVVRCCLAENSMG--VSTKAELRV 5216  
 Db 1083 MKNKMEIREDPKFLIT-NYQGVLTUNIRPFFDAGTTCRAVNELGALAECKLEVRV 1140  
 RESULT 15  
 UN52 CAEEL STANDARD; PRT: 3375 AA.  
 ID UN52 CAEEL STANDARD; Q9XTD2; Q9XTI5;  
 AC Q06561; O18263; O18263; Q9XTD2; Q9XTI5;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Basement membrane proteoglycan precursor (Perlecan homolog)  
 DE (Uncoordinated protein 52).  
 GN UNC-52 OR ZC101.2  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 [1]  
 RP SEQUENCE FROM N.A. (ISOFORM A), AND FUNCTION.  
 RX MEDLINE=93339574; PubMed=8393416;  
 RA Rogalski T.M., Williams B.D., Mullen G.P., Moerman D.G.;  
 RT "Products of the unc-52 gene in Caenorhabditis elegans are homologous  
 RT to the core protein of the mammalian basement membrane heparan  
 RT sulfate proteoglycan.";  
 RL Genes Dev. 7:1471-1484(1993).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Percy C.M., Baynes C.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RN REVISIONS, AND ALTERNATIVE SPLICING.  
 RP Durbin R.;  
 RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 RL -1- FUNCTION: Probable role in myofilament assembly and/or attachment  
 CC of the myofilament lattice to the cell membrane. May be an  
 CC extracellular anchor for integrin receptors in muscle.  
 CC -1- SUBCELLULAR LOCATION: Extracellular matrix.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=alternative splicing; Named isoforms=4;  
 CC Name=e;  
 CC IsoId=Q06561-1; Sequence=Displayed;  
 CC Note=No experimental confirmation available;  
 CC Name=a;  
 CC IsoId=Q06561-2; Sequence=VSP\_007195, VSP\_007196;  
 CC Name=b;  
 CC IsoId=Q06561-3; Sequence=VSP\_007191, VSP\_007192;  
 CC Note=No experimental confirmation available;  
 CC Name=c;  
 CC IsoId=Q06561-4; Sequence=VSP\_007193, VSP\_007194, VSP\_007195,  
 CC VSP\_007196;  
 CC Note=No experimental confirmation available;  
 CC -1- TISSUE SPECIFICITY: Found in the basement membrane of all  
 CC contractile tissues. It is concentrated over muscle dense bodies  
 CC and M-lines which are associated with beta-integrin.  
 CC -1- DEVELOPMENTAL STAGE: Synthesized early in embryogenesis.  
 CC -1- SIMILARITY: Contains 3 LDL-receptor class A domains.  
 CC -1- SIMILARITY: Contains 16 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 7 laminin EGF-like domains.  
 CC -1- SIMILARITY: Contains 3 laminin G-like domains.  
 CC -1- SIMILARITY: Contains 2 laminin IV domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----

DR EMBL; L13458; AAA28156.1; -  
 DR EMBL; Z93375; CAB07567.1; -  
 DR EMBL; Z93375; CAB07567.1; JOINED.  
 DR EMBL; Z93375; CAB07568.1; -  
 DR EMBL; Z93375; CAB07568.1; JOINED.  
 DR EMBL; Z93375; CAB07569.1; -  
 DR EMBL; Z93375; CAB07569.1; JOINED.  
 DR EMBL; Z93375; CAB07569.1; JOINED.  
 DR EMBL; Z93375; CAB07704.1; -  
 DR EMBL; Z93375; CAB07704.1; -  
 DR EMBL; Z93375; CAB07706.1; -  
 DR EMBL; Z93375; CAB07706.1; JOINED.  
 DR EMBL; Z93375; CAB07707.1; -  
 DR EMBL; Z93375; CAB07707.1; JOINED.  
 DR EMBL; Z93375; CAB07708.1; -  
 DR EMBL; Z93375; CAB07708.1; JOINED.  
 DR HSP; P01130; 1LDR.  
 DR WormPep; ZC101.2a; CE15028.  
 DR WormPep; ZC101.2b; CE15030.  
 DR WormPep; ZC101.2c; CE15034.  
 DR WormPep; ZC101.2e; CE18424.  
 DR GO; GO:000578; C:extracellular matrix; IEP.  
 DR GO; GO:0030239; P:myofibril assembly; IEP.  
 DR InterPro; IP0008985; ConA\_like Lec-gl.  
 DR InterPro; IP000881; EGF\_Ca.  
 DR InterPro; IP0006209; EGF\_Like.  
 DR InterPro; IP0007110; IG-like.  
 DR InterPro; IP0003599; IG.  
 DR InterPro; IP0003598; IG\_c2.  
 DR InterPro; IP0000034; Laminin\_B.  
 DR InterPro; IP0002049; Laminin\_EGF.  
 DR InterPro; IP0001791; Laminin\_G.  
 DR InterPro; IP0002172; LDL\_receptor\_A.  
 DR Pfam; PF00047; IG; 16.  
 DR Pfam; PF00052; laminin\_B; 2.  
 DR Pfam; PF00053; laminin\_EGF; 5.  
 DR Pfam; PF00057; LDL\_recept\_a; 3.  
 DR PRINTS; PR00261; LDLRECEPTOR.  
 DR ProDom; PD003031; Laminin\_B; 2.  
 DR SMART; SM00181; EGF; 6.  
 DR SMART; SM00179; EGF\_CA; 2.  
 DR SMART; SM00180; EGF\_Lam; 6.  
 DR SMART; SM00409; IG; 17.  
 DR SMART; SM00408; IGc2; 17.  
 DR SMART; SM00381; Lamb; 1.  
 DR SMART; SM00282; Lamb; 3.  
 DR SMART; SM00192; LDLa; 3.  
 DR PROSITE; PS00022; EGF\_1; 7.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS00026; EGF\_3; 3.  
 DR PROSITE; PS00835; IG\_Like; 17.  
 DR PROSITE; PS00025; LAM\_G\_DOMAIN; 3.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 7.  
 DR PROSITE; PS01209; LDLRA\_1; 3.  
 DR PROSITE; PS00068; LDLRA\_2; 3.  
 KW Proteoglycan; Glycoprotein; Signal; Immunoglobulin domain; Repeat;  
 KW Basement membrane; Extracellular matrix; Alternative splicing;  
 KW Laminin EGF-like domain.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 3375 BASEMENT MEMBRANE PROTEOGLYCAN.  
 FT DOMAIN 45 130 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 148 184 LDL-RECEPTOR CLASS A 1.  
 FT DOMAIN 189 225 LDL-RECEPTOR CLASS A 2.  
 FT DOMAIN 232 269 LDL-RECEPTOR CLASS A 3.  
 FT DOMAIN 271 355 IG-LIKE C2-TYPE 2.  
 FT DOMAIN 384 431 LAMININ EGF-LIKE 1 (INCOMPLETE).  
 FT DOMAIN 432 441 LAMININ EGF-LIKE 2 (N-TERMINAL).

FT DOMAIN 442 633 LAMININ DOMAIN IV 1.  
 FT DOMAIN 634 666 LAMININ EGF-LIKE 2 (C-TERMINAL).  
 FT DOMAIN 671 720 LAMININ EGF-LIKE 3 (INCOMPLETE).  
 FT DOMAIN 721 730 LAMININ EGF-LIKE 4 (N-TERMINAL).  
 FT DOMAIN 731 921 LAMININ DOMAIN IV 2.  
 FT DOMAIN 922 954 LAMININ EGF-LIKE 4 (C-TERMINAL).  
 FT DOMAIN 955 1004 LAMININ EGF-LIKE 5.  
 FT DOMAIN 1011 1060 LAMININ EGF-LIKE 6.  
 FT DOMAIN 1061 1111 LAMININ EGF-LIKE 7.  
 FT DOMAIN 1126 1222 IG-LIKE C2-TYPE 3.  
 FT DOMAIN 1226 1311 IG-LIKE C2-TYPE 4.  
 FT DOMAIN 1319 1401 IG-LIKE C2-TYPE 5.  
 FT DOMAIN 1410 1499 IG-LIKE C2-TYPE 6.  
 FT DOMAIN 1503 1585 IG-LIKE C2-TYPE 7.  
 FT DOMAIN 1588 1800 IG-LIKE C2-TYPE 8.  
 FT DOMAIN 1690 1785 IG-LIKE C2-TYPE 9.  
 FT DOMAIN 1793 1878 IG-LIKE C2-TYPE 10.  
 FT DOMAIN 1886 1970 IG-LIKE C2-TYPE 11.  
 FT DOMAIN 1973 2069 IG-LIKE C2-TYPE 12.  
 FT DOMAIN 2073 2163 IG-LIKE C2-TYPE 13.  
 FT DOMAIN 2173 2260 IG-LIKE C2-TYPE 14.  
 FT DOMAIN 2263 2343 IG-LIKE C2-TYPE 15.  
 FT DOMAIN 2349 2435 IG-LIKE C2-TYPE 16.  
 FT DOMAIN 2446 2530 IG-LIKE C2-TYPE 17.  
 FT DOMAIN 2532 2713 LAMININ G-LIKE 1.  
 FT DOMAIN 2793 2960 LAMININ G-LIKE 2.  
 FT DOMAIN 2961 3093 GLU-RICH.  
 FT DOMAIN 2972 3066 THR-RICH.  
 FT DOMAIN 3180 3359 LAMININ G-LIKE 3.  
 FT DISULFID 66 114 BY SIMILARITY.  
 FT DISULFID 149 161 BY SIMILARITY.  
 FT DISULFID 156 174 BY SIMILARITY.  
 FT DISULFID 168 183 BY SIMILARITY.  
 FT DISULFID 190 202 BY SIMILARITY.  
 FT DISULFID 197 215 BY SIMILARITY.  
 FT DISULFID 209 224 BY SIMILARITY.  
 FT DISULFID 233 246 BY SIMILARITY.  
 FT DISULFID 240 259 BY SIMILARITY.  
 FT DISULFID 253 268 BY SIMILARITY.  
 FT DISULFID 955 964 BY SIMILARITY.  
 FT DISULFID 957 971 BY SIMILARITY.  
 FT DISULFID 974 983 BY SIMILARITY.  
 FT DISULFID 986 1002 BY SIMILARITY.  
 FT DISULFID 1011 1021 BY SIMILARITY.  
 FT DISULFID 1013 1027 BY SIMILARITY.  
 FT DISULFID 1030 1039 BY SIMILARITY.  
 FT DISULFID 1042 1058 BY SIMILARITY.  
 FT DISULFID 1061 1069 BY SIMILARITY.  
 FT DISULFID 1063 1079 BY SIMILARITY.  
 FT DISULFID 1082 1091 BY SIMILARITY.  
 FT DISULFID 1094 1109 BY SIMILARITY.  
 FT DISULFID 1152 1200 BY SIMILARITY.  
 FT DISULFID 1338 1384 BY SIMILARITY.  
 FT DISULFID 1435 1481 BY SIMILARITY.  
 FT DISULFID 1527 1573 BY SIMILARITY.  
 FT DISULFID 1618 1663 BY SIMILARITY.  
 FT DISULFID 1719 1767 BY SIMILARITY.  
 FT DISULFID 1814 1861 BY SIMILARITY.  
 FT DISULFID 1907 1954 BY SIMILARITY.  
 FT DISULFID 2038 2053 BY SIMILARITY.  
 FT DISULFID 2099 2147 BY SIMILARITY.  
 FT DISULFID 2195 2242 BY SIMILARITY.  
 FT DISULFID 2284 2329 BY SIMILARITY.  
 FT DISULFID 2374 2420 BY SIMILARITY.  
 FT DISULFID 2467 2514 BY SIMILARITY.  
 FT CARBOHYD 1422 1422 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2476 2476 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2950 2950 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 3143 3143 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 1.3%; Score 519; DB 1; Length 3375;  
 Best Local Similarity 20.0%; Fred. No. 4.8e-10;  
 Matches 709; Conservative 415; Mismatches 1321; Indels 1108; Gaps 185;





Db 1927 LPHDVTLEPNILDFNSGR-----SELNGDYTCCTASNPIGEASDHGNVNI----- 1971  
QY 1976 SSSAALRVREAPVLFKKLEP-----QTVBERSSTVLEV--ELTRPWPPELRWTRNATALA 2028  
Db 1972 -----GPSLTVKTNPGPKLIVTGEPLQVKEAFCAPOCPPEVEWNLHD----- 2016  
QY 2029 PG-----KNVEIHAEGARHLVHNVPFADRGFFGCEPDDKTQAKLTVEMRQVR 2078  
Db 2017 PGPBGLDPPDFKPVITSEQPIRH-----PNVGLNAGVYTKGSSAHATATKMIYIEWE 2072  
QY 2079 LVR-----GLQAVEAREQG-----TAT---MEVOLSHADVDGSMTRGLRFQOGPTC 2122  
Db 2073 PSRTATVSLGGSSQWFDQEGKELICTATGSSLVDRLEWEKVDQDQLPDVEEHNFPGLL 2132  
QY 2123 HL-----AVRGMHTLTLGLRPEDSGLMVFKAEGVHTSARLVVTELPLVFSRPLQDVYT 2177  
Db 2133 HFPSPKNSYAGEYR---CNGYR--NNEIIASAAVHVHSSAN--ADDEPKVEIEBPPRVVV 2185  
QY 2178 TEKEKVTLECEL--SRPNVDVRM--LXQGV---ELRAGKTMATAAQACRSLTIVRCEFA 2230  
Db 2186 SQGNIVLKCQVQAEHGFKALLRGSLVRQLGTEPTEITKADPPSNDPFGYRNV 2245  
QY 2231 DQGVVCDADHAQSSAVKQGRYTYLIRYRLVLAEDAGEIQFVAENAESRAQLRVKELPV 2290  
Db 2246 DNGLVIGS--AFTAVSVGOQDKSHAQI---VKFDDKSDASFT----- 2283  
QY 2291 TLVRPLRDKIAMEKHGVLQCV--SRASAQVRHFKSQBELQPGPKVELYSDGLYRKLIIIS 2349  
Db 2284 -----CPIYVFGSKVDMTYENGDLPS---KAVPNG--NKIEIK 2317  
QY 2350 DVHAEDEDTYTC-----DAGDVK--TSAQFFVEQSIIVRGLQDVTVMEPA-----PAWFEC 2400  
Db 2318 EFDDASAGTVYCKVSFGNVVGEFVTAQMFVPD---TIIQVLLVSESSESQIGDRAWFDC 2374  
QY 2401 E-TSIPSV-----RPPKLLGKTVLQAGNVGL-----EQGTVHRLMLRRTCTST 2444  
Db 2375 KVTGDSAVISWTKEGNDLPP-----NAQVTGGRLLFTDLKEDNAGVYRCVAK---T 2424  
QY 2445 MTGPVHF-TV-----GKSRSSARLWSDIPVWLTRPLEPKTRELQSVLSCDFRPAPK- 2497  
Db 2425 KAGPLQTFVTLNVGSGQDQVTFVADSLPVVT-----VG---QPAYLSCLCKTETXP 2475  
QY 2498 --AVQWYKDDTPLSPSEKFKXSLGQMAELRILRLMPADAGVYRCQAGSAHSSTEVVEA 2555  
Db 2476 NQSVVTKKEGDL-PS-----GSRVEQGVLM-----PSVHRDDE----- 2509  
QY 2556 REVTVTGLQDABATEEGWASFCSELSHDEEVEWSLNGMPLYNDSFHEISHKGRHRLV 2615  
Db 2510 -----GSYCEIIVKEENPV---FSVVDLQIDDFIPV----- 2537  
QY 2616 LKSIQRADAGIVRASSLKVSTARSLEVRVYVFLKALDDLSAEBERTIALQ--CEVSDP 2673  
Db 2538 -----IDGEP1-----ELPPLSDEEIVNLDIEITLNTANP 2567  
QY 2674 EAHVVRKDGVOQLP--SDKYDFLHTAGTGLVHVDSVPEDAG--LYTCHVGS-----BET 2725  
Db 2568 KG-IIFETKRINGDGLLATYDTH-----HEAKTDYGTVLYEFDIGNGROIET 2617  
QY 2726 RARVVRHDLHVGTIKRLKTMVLEGESCSPECVLSHESAD-----PAMTTVGKGTVG 2778  
Db 2618 TNPINPNEWNVIKNDKNQVTIQ-----LNDESATIRQHTNPLPSLSTGVNRPVF 2668  
QY 2779 SSSRFQATROGRKVLVVRREAPSDAGEVVSFVSGTASLIVRERPAALIKPLEDQWV 2838  
Db 2669 IGRHEPTNEANDP-----RGIIIS-----QVY 2690  
QY 2839 APGEDVELRCSELGRAGTFVHVLKDKRAIKRSQKYDVVCEGTAMLVIRGASILKDA----- 2893  
Db 2691 LSGHNV-----GLGDARIPSSVWXYD-ACAST--NLCLNGANCENANNHG 2733  
QY 2894 -----GEY-----TCEVEASKSTASLHVEBKANCFTB-ELTNLQVBEKGTAV-- 2934

Db 2734 FSCCAEERFHGEYQOWRNSNSCHDESCNTGICLDNEESQWCVCPLGTTGLRCEEK-TEIPQ 2792  
QY 2935 ---FTCKTEHPAA--TVTWKGLLELRA---SGKH-----OPSQ 2965  
Db 2793 PLGFTSDTSFLAVKEPVKFESIKMKLRPQADSDEHILMYFASDYGSNTKQYTSLSLIANQ 2852  
QY 2966 EGLTLRLIISALEKADSDTYTCD-----IGQASRAQLLVQGRVHIELEDVDVQEG 3019  
Db 2853 VVLTVRRPDKEVQKIRSETLEAGELIDVAVRQAGNALVMTVDGNQVSTIE---TDTLKP 2908  
QY 3020 SSATERCRISPNYEPVHWFLDKTPHLANELNEIDAQPGGYHVLTLRO---LALKDSGT 3075  
Db 2909 GTEIFIGLPGCLNSP-----DDVVEQSFOGCVVEIILNSQDVLQNLSSGD 2956  
QY 3076 IY-----FEAGQORASAAIR---VTERKPSVFSRELTDATITE--GEDTLVCTETST 3121  
Db 2957 ISSCEESQFPVEEDDTTITTTTTEPPEAVIERPTEETPTTEETPTTEETPTTEETPTT 3016  
QY 3122 CDIPMCWTCKGKT 3134  
Db 3017 TEEPTTTTEETPT 3029

Search completed: September 13, 2004, 11:25:53  
Job time : 127 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 11:12:23 ; Search time 461 Seconds  
(without alignments)  
5453.469 Million cell updates/sec

Title: US-10-077-130-5  
Perfect score: 41273  
Sequence: 1 MDPQSGAPFLTRPKAFV.....RNREKRALLYKHNLAQVR 7968

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32134	77.9	6620	4 Q96AA2	Q96aa2 homo sapien
2	8486.5	20.6	2212	4 Q8NH3	Q8nhn3 homo sapien
3	6302	16.7	1319	4 Q8HCD3	Q8hcd3 homo sapien
4	6619.5	16.0	1596	4 Q8HCL6	Q8hcl6 homo sapien
5	5358.5	13.0	1040	4 Q8NH2	Q8nhn2 homo sapien
6	3865	9.4	767	4 Q8GYC7	Q8gyc7 homo sapien
7	3580.5	8.7	26926	4 Q8WZB3	Q8wzb3 homo sapien
8	3577.5	8.7	26926	4 Q10466	Q10466 homo sapien
9	3427	8.3	34350	4 Q8WZ42	Q8wz42 homo sapien
10	3389	8.2	658	4 Q8NH4	Q8nhn4 homo sapien
11	3313	8.0	7962	4 Q10465	Q10465 homo sapien
12	3270	7.9	1020	4 Q8NH0	Q8nhn0 homo sapien
13	3182.5	7.7	8081	5 Q7Z120	Q7z120 caenorhabdi
14	3113	7.5	646	4 Q8NH8	Q8nhn8 homo sapien
15	2793	6.8	551	4 Q8NH7	Q8nhn7 homo sapien
16	2653	6.4	542	4 Q8NH5	Q8nhn5 homo sapien

17	2629	6.4	4162	13 Q98918	Q98918 gallus gall
18	2392.5	5.8	4076	11 Q7TN00	Q7tn00 rattus norv
19	2291	5.6	4039	13 Q7ZZ46	Q7zz46 brachydanio
20	2273	5.5	5516	13 Q7ZZ48	Q7zz48 brachydanio
21	2203	5.3	3262	11 Q9EQJ5	Q9eqj5 mus musculu
22	2189.5	5.3	5604	4 Q8WZ53	Q8wz53 homo sapien
23	2168	5.3	3950	6 Q7YRF5	Q7yrf5 canis famil
24	2087	5.1	16215	5 Q8NFS3	Q8nfs3 drosophila
25	2087	5.1	18074	5 Q917U4	Q917u4 drosophila
26	2072.5	5.0	5636	4 Q96RW7	Q96rw7 homo sapien
27	2067	5.0	2242	4 Q96RP9	Q96rp9 homo sapien
28	2048	5.0	5198	5 Q76518	Q76518 caenorhabdi
29	2043.5	5.0	5175	5 Q810L3	Q810l3 caenorhabdi
30	1996.5	4.8	803	11 Q80ZP5	Q80zf5 rattus norv
31	1956.5	4.7	17352	5 Q85VM2	Q85ym2 procambarus
32	1896	4.6	4816	5 Q8T103	Q8t103 bombyx mori
33	1881.5	4.6	8625	5 Q86GD6	Q86gd6 procambarus
34	1837	4.5	4824	5 Q95YM1	Q95ym1 procambarus
35	1763.5	4.3	4796	5 Q9W055	Q9w055 drosophila
36	1736.5	4.2	4796	5 Q9NL88	Q9nl88 drosophila
37	1677	4.1	8943	5 Q9V4F7	Q9v4f7 drosophila
38	1665	4.0	18519	5 Q8ISF6	Q8isf6 caenorhabdi
39	1664.5	4.0	18534	5 Q8ISF7	Q8isf7 caenorhabdi
40	1650.5	4.0	7158	5 Q23551	Q23551 caenorhabdi
41	1635	4.0	9270	5 Q8MLD9	Q8mld9 drosophila
42	1619.5	3.9	19066	13 Q801W8	Q801w8 brachydanio
43	1619	3.9	6831	5 Q23550	Q23550 caenorhabdi
44	1612	3.9	345	11 Q811H7	Q811h7 mus musculu
45	1557.5	3.8	4736	5 Q7YT99	Q7yt99 mytilus gal

## ALIGNMENTS

## RESULT 1

Q96AA2	PRELIMINARY;	PRT; 6620 AA.
ID	Q96AA2	
AC	Q96AA2;	
DT	01-DEC-2001 (TrEMBLrel. 19, Created)	
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	Obscurin.	
GN	OBSCN.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Heart;	
RX	MEDLINE=21342081; PubMed=11448995;	
RA	Young P.W., Ehler E., Gautel M.;	
RT	"Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor	
RT	protein involved in sarcomere assembly.";	
RL	J. Cell Biol. 154:123-136(2001).	
CC	-1- SIMILARITY: CONTAINS 1 PH DOMAIN.	
DR	EMBL; AJ002535; CAC44768.1; -.	
DR	Genew; HGNC:15719; OBSCN.	
DR	GO; GO:0005524; F:ATP binding; IEA.	
DR	GO; GO:0004812; F:RNA ligase activity; IEA.	
DR	GO; GO:0006418; P:amino acid activation; IEA.	
DR	InterPro; IPR003961; FN_III.	
DR	InterPro; IPR008957; FN_III-like.	
DR	InterPro; IPR007110; Ig-like.	
DR	InterPro; IPR003598; Ig_c2.	
DR	InterPro; IPR000048; IQ-region.	
DR	InterPro; IPR001849; PH.	
DR	InterPro; IPR000219; RhoGEF.	
DR	InterPro; IPR001452; SH3.	
DR	InterPro; IPR001412; tRNA-synt_1.	
DR	Pfam; PF00041; In3; 2.	
DR	Pfam; PF00047; Ig; 49.	
DR	Pfam; PF00612; IQ; 1.	



QY 1921 VVSPSDVAVVFRDQALLOPSEKFAISQSGASHSLT:SDLYLEADAGQITVEAEGASSAA 1980  
DB 1921 VVSPSDVAVVFRDQALLOPSEKFAISQSGASHSLT:SDLYLEADAGQITVEAEGASSAA 1980  
QY 1981 LRVRAPVLFKKLEPOTVEERSVTVLEVLTRPWPBLRWRNATAPAGNVEIHAEGA 2040  
DB 1981 LRVRAPVLFKKLEPOTVEERSVTVLEVLTRPWPBLRWRNATAPAGNVEIHAEGA 2040  
QY 2041 RHRLVHNVGFADRGFFCETPDPTDKTQAKLVEMRQVRLVRLGQAVAREQGTATMEVQL 2100  
DB 2041 RHRLVHNVGFADRGFFCETPDPTDKTQAKLVEMRQVRLVRLGQAVAREQGTATMEVQL 2100  
QY 2101 SHADVDSWTRDGLRFQOGPTCHLAVRGPMTLTLISGLRPEDSGLMVFKABGHTSARLV 2160  
DB 2101 SHADVDSWTRDGLRFQOGPTCHLAVRGPMTLTLISGLRPEDSGLMVFKABGHTSARLV 2160  
QY 2161 VTLPVSPSRPLQDVVITTEKEKVTLECELSRPNVDVRLKDGVELRAGKTMIAAQAQACR 2220  
DB 2161 VTLPVSPSRPLQDVVITTEKEKVTLECELSRPNVDVRLKDGVELRAGKTMIAAQAQACR 2220  
QY 2221 SLTYRCFADQGVVCDHAQAQSSASVKVQGRYTYLLIYRVLAEADAGEIQFVAENAESR 2280  
DB 2221 SLTYRCFADQGVVCDHAQAQSSASVKVQGRYTYLLIYRVLAEADAGEIQFVAENAESR 2280  
QY 2281 AOLRVKELPVLVRLRDKIAMEXHGVLECOVSRASAOVRFWFGSOBLOPQPKYELVSD 2340  
DB 2281 AOLRVKELPVLVRLRDKIAMEXHGVLECOVSRASAOVRFWFGSOBLOPQPKYELVSD 2340  
QY 2341 GLYRKLIIISDVAEDEDYTCADGVKTSQAQFVVEQSITIVRGLQDVTVMEPAPAMFEC 2400  
DB 2341 GLYRKLIIISDVAEDEDYTCADGVKTSQAQFVVEQSITIVRGLQDVTVMEPAPAMFEC 2400  
QY 2401 ETSIPSVRPPKWLIGKTVLQAGNVGLBOEGTVHRLMLRRTCSMTGPHVFTVCKSRSSA 2460  
DB 2401 ETSIPSVRPPKWLIGKTVLQAGNVGLBOEGTVHRLMLRRTCSMTGPHVFTVCKSRSSA 2460  
QY 2461 RLWVSDIIPVLTTRPLEKTRGELQSVLSCDFRPAKAVQWYKDDTPLSPSEKFKMSLEG 2520  
DB 2461 RLWVSDIIPVLTTRPLEKTRGELQSVLSCDFRPAKAVQWYKDDTPLSPSEKFKMSLEG 2520  
QY 2521 QMAELRIILRLMPADAGVTRCOAGSAHSSTEVTVAREVTVTGPIQDAEATEEGWASFSCE 2580  
DB 2521 QMAELRIILRLMPADAGVTRCOAGSAHSSTEVTVAREVTVTGPIQDAEATEEGWASFSCE 2580  
QY 2581 LSHDEEVEWSLNGMPLVNDSPHEISHKGRHTLVLSIQIRADAGIVRASSLKVSTSARL 2640  
DB 2581 LSHDEEVEWSLNGMPLVNDSPHEISHKGRHTLVLSIQIRADAGIVRASSLKVSTSARL 2640  
QY 2641 EVVRKPVVFLKALDLSAEERGTALQCEVSDPEAHVVRKDXGVQLQPSDKYDFLHTAGT 2700  
DB 2641 EVVRKPVVFLKALDLSAEERGTALQCEVSDPEAHVVRKDXGVQLQPSDKYDFLHTAGT 2700  
QY 2701 RGLVVDVSPDAGLYTCHVGSEETRAVRVHDLHVGITKELKTMEVLEGESCSFECVLS 2760  
DB 2701 RGLVVDVSPDAGLYTCHVGSEETRAVRVHDLHVGITKELKTMEVLEGESCSFECVLS 2760  
QY 2761 HESASDPAMWTVGGKTVGSSSRFQATROGRKYLIVREAPSDADAGEVVSFVRLGTSKASL 2820  
DB 2761 HESASDPAMWTVGGKTVGSSSRFQATROGRKYLIVREAPSDADAGEVVSFVRLGTSKASL 2820  
QY 2821 IVREPAIILKPLEDQWAPGEDVELRCELSRAGTPVHVLKORKAIRKSQKYDVVVCBGTM 2880  
DB 2821 IVREPAIILKPLEDQWAPGEDVELRCELSRAGTPVHVLKORKAIRKSQKYDVVVCBGTM 2880  
QY 2881 AMLVIRGASLKDAGEYTCVEBAKSTASLHVEEKANCFTBELTNLQVEEKGTAVFTCKTE 2940  
DB 2881 AMLVIRGASLKDAGEYTCVEBAKSTASLHVEEKANCFTBELTNLQVEEKGTAVFTCKTE 2940  
QY 2941 HPAATVWRKGLLELRASGKHQPSOEGTLRLITISALEKADSDYTCDIGAQASRAQLIV 3000  
DB 2941 HPAATVWRKGLLELRASGKHQPSOEGTLRLITISALEKADSDYTCDIGAQASRAQLIV 3000

QY 3001 QQRVHHIIEBLEDVQVEGSSATFRCRISPNYEPVHWFLDKTPHLANELNIDAOPGGY 3060  
DB 3001 QQRVHHIIEBLEDVQVEGSSATFRCRISPNYEPVHWFLDKTPHLANELNIDAOPGGY 3060  
QY 3061 HVLTLRQLAKDSGTYIFEAGDORASAAALRVTEKESVFSRELTDTATIEGEDTLVCETS 3120  
DB 3061 HVLTLRQLAKDSGTYIFEAGDORASAAALRVTEKESVFSRELTDTATIEGEDTLVCETS 3120  
QY 3121 TCDIPMCWTQDKTILRGSRARCOLSHEGHRAQLIITGATLQDSGRYKCEAGGACSSIVRV 3180  
DB 3121 TCDIPMCWTQDKTILRGSRARCOLSHEGHRAQLIITGATLQDSGRYKCEAGGACSSIVRV 3180  
QY 3181 HARVRFQALKDLEVLGGAATLRCVLSVAAPVKWCYGNVLPDGRKYSLRQEGAMLE 3240  
DB 3181 HARVRFQALKDLEVLGGAATLRCVLSVAAPVKWCYGNVLPDGRKYSLRQEGAMLE 3240  
QY 3241 LVVRNLRPDSGRYSCSPDQTTSATLTVTALPAQFIGLKRNEKATEGATATLRCELSKT 3300  
DB 3241 LVVRNLRPDSGRYSCSPDQTTSATLTVTALPAQFIGLKRNEKATEGATATLRCELSKT 3300  
QY 3301 APVEMKGSSETLDRGDRYCLRODGMCELOIISGLAVMDAAEYSCVCGEBERTSASLTIRPM 3360  
DB 3301 APVEMKGSSETLDRGDRYCLRODGMCELOIISGLAVMDAAEYSCVCGEBERTSASLTIRPM 3360  
QY 3361 PAHFIGELRHQBSIEGATATLRCELSKAAPVEMKRGRESLRDGRHSLRQDGAVALCQIC 3420  
DB 3361 PAHFIGELRHQBSIEGATATLRCELSKAAPVEMKRGRESLRDGRHSLRQDGAVALCQIC 3420  
QY 3421 GLAVADAGSYSCVCGEBERTSASLTIVKALPAKTEGLRNEEAVEGATAMLCESKVAPVE 3480  
DB 3421 GLAVADAGSYSCVCGEBERTSASLTIVKALPAKTEGLRNEEAVEGATAMLCESKVAPVE 3480  
QY 3481 WRKGPENLRDGRYILRQEGTRCELOICGLAVADAGEYLCVCGQERTSATLTIRALPARF 3540  
DB 3481 WRKGPENLRDGRYILRQEGTRCELOICGLAVADAGEYLCVCGQERTSATLTIRALPARF 3540  
QY 3541 IEDVKQNEAREGATAVLQCELSKAAPVEMKRGSETLRDGRYSRLQDGTKCELIQIRGLAM 3600  
DB 3541 IEDVKQNEAREGATAVLQCELSKAAPVEMKRGSETLRDGRYSRLQDGTKCELIQIRGLAM 3600  
QY 3601 ADTGEYSVCVCGQERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCELSKVAPVEMWK 3660  
DB 3601 ADTGEYSVCVCGQERTSAMLTVRALPIKFTTEGLRNEEATEGATAVLRCELSKVAPVEMWK 3660  
QY 3661 HETLRDGRHSLRQDGAVALCQICGLAVADAGEYLCVCGQERTSAMLTVRAMPSKFIEGL 3720  
DB 3661 HETLRDGRHSLRQDGAVALCQICGLAVADAGEYLCVCGQERTSAMLTVRAMPSKFIEGL 3720  
QY 3721 RNEEATEGATATLRCELSKAAPVEMKRGHETLRDGRHSLRQDGSCELIQIRGLAVVDAG 3780  
DB 3721 RNEEATEGATATLRCELSKAAPVEMKRGHETLRDGRHSLRQDGSCELIQIRGLAVVDAG 3780  
QY 3781 EYSCVCGQERTSATLTIRALPARFIEDVKNQAREGATAVLQCELSKAAPVEMKRGSETL 3840  
DB 3781 EYSCVCGQERTSATLTIRALPARFIEDVKNQAREGATAVLQCELSKAAPVEMKRGSETL 3840  
QY 3841 RGGDRYSRLQDGTTRCELOIHGLSVADTGEYSVCVCGQERTSATLTVPAPQVPREPLOSLO 3900  
DB 3841 RGGDRYSRLQDGTTRCELOIHGLSVADTGEYSVCVCGQERTSATLTVPAPQVPREPLOSLO 3900  
QY 3901 AEBGSTATLQCELSSEPTATVWMSKGLQLOQANGREPRLOQGTAEVLQDLOREDTGEYT 3960  
DB 3901 AEBGSTATLQCELSSEPTATVWMSKGLQLOQANGREPRLOQGTAEVLQDLOREDTGEYT 3960  
QY 3961 CTCGSQATSATLTVTAAAPVRFLRELOHQVEDEGGTAHLCCELSRAGASVEMKRGSLQLPP 4020  
DB 3961 CTCGSQATSATLTVTAAAPVRFLRELOHQVEDEGGTAHLCCELSRAGASVEMKRGSLQLPP 4020  
QY 4021 CAKYQMVQGAALVVRGVEQEDAGDYTCDTGHTOSMASLSVRVPRPKFKTRQLQLEOB 4080  
DB 4021 CAKYQMVQGAALVVRGVEQEDAGDYTCDTGHTOSMASLSVRVPRPKFKTRQLQLEOB 4080  
QY 4081 TGDIALCCQLSDAESGAVVQMLKEGVHLAGPKYEMRSGQATRELLIHLQEAKTGEYA 4140

4081 TGDARLCCQLSDAESGAVVOWLKEGVHLAGPKYEMRSQATRELLIHQLEAKDTGEYA 4140  
4141 CVTGQKTAASLRVTEPEVTIVRGLVDAEVTADSEVSESVSAGATGVQWCLQGLPLQ 4200  
4141 CVTGQKTAASLRVTEPEVTIVRGLVDAEVTADSEVSESVSAGATGVQWCLQGLPLQ 4200  
4201 SNEVTEAVRGRIRHTLRKGVTPEDAGTVSFHGNHASSAQLTVRAPEVTILEPQDVQ 4260  
4201 SNEVTEAVRGRIRHTLRKGVTPEDAGTVSFHGNHASSAQLTVRAPEVTILEPQDVQ 4260  
4261 LSEGQDASFCQLSRASQGEARWALGGVPLQANEMNDITVEQGTLLHLTKHVKLTLEDAGT 4320  
4261 LSEGQDASFCQLSRASQGEARWALGGVPLQANEMNDITVEQGTLLHLTKHVKLTLEDAGT 4320  
4321 VSFHVTGTCSSAQLKVTAKNTVVRGLENVVALEGEALFECOLSQPEVAHTWLLDDEPV 4380  
4321 VSFHVTGTCSSAQLKVTAKNTVVRGLENVVALEGEALFECOLSQPEVAHTWLLDDEPV 4380  
4381 RTSNAEAVVFFENGRLHLLLNKLRPQDSCHVTVFLAGDMVTSAPLTVRGMRLLEPLKN 4440  
4381 RTSNAEAVVFFENGRLHLLLNKLRPQDSCHVTVFLAGDMVTSAPLTVRGMRLLEPLKN 4440  
4441 AAVRAGAQAARTCTLSEAVPVGBASWYINGAAVOPDDSDMTVTRADGSHQALLRSAPHH 4500  
4441 AAVRAGAQAARTCTLSEAVPVGBASWYINGAAVOPDDSDMTVTRADGSHQALLRSAPHH 4500  
4501 AGEVTFACRDVAVASARLVGLPPEDEAVVAHSSHTVLSWAAPMSDGGGLCGYRVE 4560  
4501 AGEVTFACRDVAVASARLVGLPPEDEAVVAHSSHTVLSWAAPMSDGGGLCGYRVE 4560  
4561 VKEGATQWRLCHELVPGPECVDPGLAPGETYFRVAAVGPVGAEPVHLPTQVRLAEP 4620  
4561 VKEGATQWRLCHELVPGPECVDPGLAPGETYFRVAAVGPVGAEPVHLPTQVRLAEP 4620  
4621 KPVPQPSAPSRQVAGEDVSLVVAEAGEVIVHKGMRLOPGRFVSVQGRQOML 4680  
4621 KPVPQPSAPSRQVAGEDVSLVVAEAGEVIVHKGMRLOPGRFVSVQGRQOML 4680  
4681 VIKGFTABDQGEYHCGLAQGSICPAATFQVALSPASVDEAPQPSLPPEAAQEGDLHLW 4740  
4681 VIKGFTABDQGEYHCGLAQGSICPAATFQVALSPASVDEAPQPSLPPEAAQEGDLHLW 4740  
4741 BALARKRMSREPTLDSISELPEEDGRSQRLOPQABEVAVDLSEGYSTADELARTGDADL 4800  
4741 BALARKRMSREPTLDSISELPEEDGRSQRLOPQABEVAVDLSEGYSTADELARTGDADL 4800  
4801 SHTSSDDESBRAGTSLVTLKACRPGTSPILASKVGAAPAPSVKPOQOEPLAAVRPPLG 4860  
4801 SHTSSDDESBRAGTSLVTLKACRPGTSPILASKVGAAPAPSVKPOQOEPLAAVRPPLG 4860  
4861 DLSTKDLGDPNDKAAVKIQAAPKGYKVRKEMKQEGPMFSTFTGDEAQVGDALRLCV 4920  
4861 DLSTKDLGDPNDKAAVKIQAAPKGYKVRKEMKQEGPMFSTFTGDEAQVGDALRLCV 4920  
4921 VASKADVRAWLKDGVELTDCGRHHIDQLGDGTCSLLIAGLDRADAGCYTCQVSNKFGOV 4980  
4921 VASKADVRAWLKDGVELTDCGRHHIDQLGDGTCSLLIAGLDRADAGCYTCQVSNKFGOV 4980  
4981 THSACVVVSGSESAESSGGELDDAFRAARLRHLFRTKSPAESVDEELFSLADEGPA 5040  
4981 THSACVVVSGSESAESSGGELDDAFRAARLRHLFRTKSPAESVDEELFSLADEGPA 5040  
5041 EPPEPADQWTVREDEHFCIRFEALTARQAVTRFQEMFATLGIIVEIKLVEQGRVEM 5100  
5041 EPPEPADQWTVREDEHFCIRFEALTARQAVTRFQEMFATLGIIVEIKLVEQGRVEM 5100  
5101 CISKETPAPVPPPEPLSLTSDAAPVFLTELQNEQVQDGYPSVFCVWTGQPMPSVRWF 5160  
5101 CISKETPAPVPPPEPLSLTSDAAPVFLTELQNEQVQDGYPSVFCVWTGQPMPSVRWF 5160  
5161 KDGKLEEDDHYMTNEDQGGHOLIITAVVPADMGVYRCLAESMGVSSTKAELRVDLTS 5220

5161 KDGKLEEDDHYMTNEDQGGHOLIITAVVPADMGVYRCLAESMGVSSTKAELRVDLTS 5220  
5221 TDYDTAADAATESSSYFSAQGYLSSREQEGTSTTDEGQLPQVVEBELRLQVAPGTRLAKF 5280  
5221 TDYDTAADAATESSSYFSAQGYLSSREQEGTSTTDEGQLPQVVEBELRLQVAPGTRLAKF 5280  
5281 QLKVXGYPAPRPLYWFKOGQPLTASAHIRMTGKILHLEIIISVTREDSGQAAVTSNANG 5340  
5281 QLKVXGYPAPRPLYWFKOGQPLTASAHIRMTGKILHLEIIISVTREDSGQAAVTSNANG 5340  
5341 AAYSSARLLVRGPDPPEEKPSADVHEQLVPPRMLERFTPKVKKGSSITFSVKVEGRVPP 5400  
5341 AAYSSARLLVRGPDPPEEKPSADVHEQLVPPRMLERFTPKVKKGSSITFSVKVEGRVPP 5400  
5401 TVHMLREAEAGVLIWIGPDTTCYTVASSAQOHSVLVDVGRHOGTYTCTIASNAGQALC 5460  
5401 TVHMLREAEAGVLIWIGPDTTCYTVASSAQOHSVLVDVGRHOGTYTCTIASNAGQALC 5460  
5461 SASLHVSGLPKVEQEKEKKEALISTFLOGTTQALSAQGLETFADLGGQKKEPPLAAKE 5520  
5461 SASLHVSGLPKVEQEKEKKEALISTFLOGTTQALSAQGLETFADLGGQKKEPPLAAKE 5520  
5521 ALGHLSLAEGVTEPEFLQKLTQITQITMVSATKTOAKLQVPGGSDSDSKTPSASPRHGRSR 5580  
5521 ALGHLSLAEGVTEPEFLQKLTQITQITMVSATKTOAKLQVPGGSDSDSKTPSASPRHGRSR 5580  
5581 PSSSIQESSSESDGDARGEIFDIYVVTADYLPKAEQDAITLREGQVVEVLDAAHPLRW 5640  
5581 PSSSIQESSSESDGDARGEIFDIYVVTADYLPKAEQDAITLREGQVVEVLDAAHPLRW 5640  
5641 LVRTPTKSSPSRQGWSPAYLDRRLKLSPEWGAABAEPEFGEAVSEDEYKARLSVIOE 5700  
5641 LVRTPTKSSPSRQGWSPAYLDRRLKLSPEWGAABAEPEFGEAVSEDEYKARLSVIOE 5700  
5701 LLSSEQAQVVEELOFQSHLOHLERCPHVPITAVAGQKAVIFRNVRDITGRPHSSFLQELQ 5760  
5701 LLSSEQAQVVEELOFQSHLOHLERCPHVPITAVAGQKAVIFRNVRDITGRPHSSFLQELQ 5760  
5761 CDTDDVAVACFTKQAAFEQYLEFLVGRVQAESVWVSTAIQEFYKKAEEALLAGDPSQP 5820  
5761 CDTDDVAVACFTKQAAFEQYLEFLVGRVQAESVWVSTAIQEFYKKAEEALLAGDPSQP 5820  
5821 PPPLOHLYLEQVVERVQYQALLKELIRNKARNQNCALLBQAVVVSALPQRAENKLHV 5880  
5821 PPPLOHLYLEQVVERVQYQALLKELIRNKARNQNCALLBQAVVVSALPQRAENKLHV 5880  
5881 SLMENYPTCTLEALGEPRIQGHFIIVMEGAPGARMKGNHNVFLFRNHLVICKPRRSRT 5940  
5881 SLMENYPTCTLEALGEPRIQGHFIIVMEGAPGARMKGNHNVFLFRNHLVICKPRRSRT 5940  
5941 DTVSVVRNMMKLSIDLDNDQVGGDDRAFEVYQBEREDSVRKYLLQARTAIKSSWKEIC 6000  
5941 DTVSVVRNMMKLSIDLDNDQVGGDDRAFEVYQBEREDSVRKYLLQARTAIKSSWKEIC 6000  
6001 GIQORLALPVWRPDPFEELADCTAEALGETVKLACRVTGTPKPIVSWYKOGKAVQVDPHH 6060  
6001 GIQORLALPVWRPDPFEELADCTAEALGETVKLACRVTGTPKPIVSWYKOGKAVQVDPHH 6060  
6061 ILIEDPDGSCALIIDSLTGVDSQVMPAASAGNCSTLGLIKLVQVPPRFNKNVRASFPV 6120  
6061 ILIEDPDGSCALIIDSLTGVDSQVMPAASAGNCSTLGLIKLVQVPPRFNKNVRASFPV 6120  
6121 EGEDAQFTCTIEGAPYPOIRWYKDGALLITGNKFTQISEPRSGLLVIRAAKSKEDLGLY 6180  
6121 EGEDAQFTCTIEGAPYPOIRWYKDGALLITGNKFTQISEPRSGLLVIRAAKSKEDLGLY 6180  
6181 ECELVNRLGSRASAEALRIQSPMLQAQOCHREOLVAAVEDTTLER 6232  
6181 ECELVNRLGSRASAEALRIQSPMLQAQOCHREOLVAAVEDTTLER 6232  
6233 SVLKELLGPKAPGSTGLTGPCCPRG-----APAL-----QBTGQOP- 6271  
6241 TVKSPRCORR-SFSKSPSRSPRCSASPLRPGLLAPOLLVLPAGQAPRAEPAGQKPV 6299

QY 6272 -PVTGTEAPAPPRV-----POPLLHEGPE-----OEPEAIARAQEWTPPIRMEGA 6317  
 Db 6300 VPTLYVTEAASHSPALGSLGPQPKWVEETIEVRVKMGPGQVSPTE--VPRSSGH 6357  
 QY 6318 AW--PGAGTG 6325  
 Db 6358 LFTLPGATPG 6367

RESULT 2  
 Q8NH3 PRELIMINARY; PRT; 2212 AA.

AC Q8NH3;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Obscurin (Fragment).  
 GN O8CN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Young P., Ehler E., Gautel M.;  
 RT "Obscurin, a giant sarcomeric Rho-GTP protein involved in sarcomere  
 assembly";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ314905; CAC85752.1; -;  
 DR InterPro; IPR003599; IG.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003598; IG\_C2.  
 DR Pfam; PF00047; IG; 24.  
 DR SMART; SM00409; IG; 25.  
 DR SMART; SM00408; IGC2; 23.  
 DR PROSITE; PS00835; IG Like; 22.  
 KW Immunoglobulin domain.  
 FT NON TER  
 FT NON TER  
 SQ SEQUENCE 2212 2212  
 B6DFA263EB749465 CRC64;

Query Match 20.6%; Score 8486.5; DB 4; Length 2212;  
 Best Local Similarity 76.5%; Pred. No. 0;  
 Matches 1692; Conservative 1; Mismatches 2; Indels 517; Gaps 3;

QY 2734 LHVGITKELKTMVELEGSCFECVLSHESASDPAMTVGKTVGSSRFOATQGRKYI 2793  
 Db 1 LHVGITKELKTMVELEGSCFECVLSHESASDPAMTVGKTVGSSRFOATQGRKYI 60

QY 2794 LVREAAPS DAGEVVFVSRGLTSKASLIVRRPAAIIPKLEDDQWVAPGEDVELRCEL SRA 2853  
 Db 61 LVREAAPS DAGEVVFVSRGLTSKASLIVRRPAAIIPKLEDDQWVAPGEDVELRCEL SRA 120

QY 2854 GTPVHNLKDKRAIRKSKYDVVCEGTMAVLIRGASLKDAGEYTCVEEASKSTASLHVEE 2913  
 Db 121 GTPVHNLKDKRAIRKSKYDVVCEGTMAVLIRGASLKDAGEYTCVEEASKSTASLHVEE 180

QY 2914 KANCFTEELTNLQVEEKGTAFTCTKTEHPAATVTVRKGLLELRASGKHQPSQEGTLRLT 2973  
 Db 181 KANCFTEELTNLQVEEKGTAFTCTKTEHPAATVTVRKGLLELRASGKHQPSQEGTLRLT 240

QY 2974 ISALEKADSDTYTCDIGQAQRAQLLVQ-GRVHHIEDLVVDVQEGSSATFRCRISPAN 3032  
 Db 241 ISALEKADSDTYTCDIGQAQRAQLLVQAGRRVHHIEDLVVDVQEGSSATFRCRISPAN 300

QY 3033 YEPVHWFELDKTLPHANELNEIDAOPGGYHVLTLROLAKDSGTIYFEAGDORASALRV 3092  
 Db 301 YEPVHWFELDKTLPHANELNEIDAOPGGYHVLTLROLAKDSGTIYFEAGDORASALRV 360

QY 3093 EKPVSFVSRRLTDATITEGEDLTLVCTSTCDIPFCMTWKDGKTLRGSARCOLSHEGHRAL 3152  
 Db 361 EKPVSFVSRRLTDATITEGEDLTLVCTSTCDIPFCMTWKDGKTLRGSARCOLSHEGHRAL 420

QY 3153 LITGATLQDSGRYKCEAGGACSSSIVRVHARPVRFOEALKDLEVLGGAAATLRCVLSVA 3212  
 Db 421 LITGATLQDSGRYKCEAGGACSSSIVRVHARPVRFOEALKDLEVLGGAAATLRCVLSVA 480

QY 3213 APVKWCYGNVLRPGDKYSLROEGAMLELVVRNLRPODSGRYSCSFGQDTTSATLVTAL 3272  
 Db 481 APVKWCYGNVLRPGDKYSLROEGAMLELVVRNLRPODSGRYSCSFGQDTTSATLVTAL 540

QY 3273 PAQFIGKLNRKEATEGATATLRCELSKTAPVWRKGSSETLRDGDYRCLRDQGMCELQIR 3332  
 Db 541 PAQFIGKLNRKEATEGATATLRCELSKTAPVWRKGSSETLRDGDYRCLRDQGMCELQIR 600

QY 3333 GLAMVDAAEYSCVCGEERTSASLTIRPMPAHFIGRLRHOESTEGATATLRCELSKAAPVE 3392  
 Db 601 GLAMVDAAEYSCVCGEERTSASLTIRPMPAHFIGRLRHOESTEGATATLRCELSKAAPVE 660

QY 3393 WRKGRESLRDGDHSLRDQGA VCELCI CGLAVADAGEYSCVCGEERTSATLTKALPAKF 3452  
 Db 661 WRKGRESLRDGDHSLRDQGA VCELCI CGLAVADAGEYSCVCGEERTSATLTKALPAKF 720

QY 3453 TEGLRNEEAVEGATAMLCELSKVAPVWRKGPENLRDGDYRILROEGTRCELCI CGLAM 3512  
 Db 721 TEGLRNEEAVEGATAMLCELSKVAPVWRKGPENLRDGDYRILROEGTRCELCI CGLAM 780

QY 3513 ADAGEYLCVCGQERTSATLITRALPARFIEDVKNQEARREGATAVLQCELNSAAPVWRKG 3572  
 Db 781 ADAGEYLCVCGQERTSATLITRALPARFIEDVKNQEARREGATAVLQCELNSAAPVWRKG 840

QY 3573 SETLRDGDYSLRDQGTKELCI RGLAMADTGEYSCVCGQERTSAMLTVRALPIKFTGEL 3632  
 Db 841 SETLRDGDYSLRDQGTKELCI RGLAMADTGEYSCVCGQERTSAMLTVRALPIKFTGEL 900

QY 3633 RNEEATEGATAVLRCEL SKVAPVWRKGHETLRDGDHSLRDQGARCELQIRGLVAEDAG 3692  
 Db 901 RNEEATEGATAVLRCEL SKVAPVWRKGHETLRDGDHSLRDQGARCELQIRGLVAEDAG 960

QY 3693 EYLCMCGKERTSAMLTVRAMP SKFIEGLRNEBATEGDTATLWCEL SKAAPVWRKGHETL 3752  
 Db 961 EYLCMCGKERTSAMLTVRAMP SKFIEGLRNEBATEGDTATLWCEL SKAAPVWRKGHETL 1020

QY 3753 RGDHSLRDQGSRCCELCI RGLAVVDAGEYSCVCGQERTSATLTVRALPARFIEDVKNQ 3812  
 Db 1021 RGDHSLRDQGSRCCELCI RGLAVVDAGEYSCVCGQERTSATLTVRALPARFIEDVKNQ 1080

QY 3813 AREGATAVLQCEL SKAAPVWRKGSSETLRGGDYSRLRDQGTCELCI RGLSVADTGEYSC 3872  
 Db 1081 AREGATAVLQCEL SKAAPVWRKGSSETLRGGDYSRLRDQGTCELCI RGLSVADTGEYSC 1140

QY 3873 VCGQERTSATLTVR----- 3886  
 Db 1141 VCGQERTSATLTVRALPARFTQDLTKEASEGATATLQCEL SKVAPVWRKGPETLRDGG 1200

QY 3887----- 3886

Db 1201 RYSLKQDGTCELCI RGLSVADAGEYSCVCGQERTSATLTVRALPARFTEGLRNEEAMEG 1260

QY 3887----- 3886

Db 1261 ATATLQCEL SKAAPVWRKGLEALRDGDKYSLRDQGA VCELCI RGLAMADNGVYSCVCG 1320

QY 3887----- 3886

Db 1321 ERTSATLTVRALPARFIEDMRNOKATEGATVTLQCLKRAAPVWRKGPNTLKDGRYSL 1380

QY 3887----- 3886

Db 1381 KQDGTSCELQIRGLVIADAGEYSCICEQERTSATLTVRALPARFIEDVRNHEATEGATAV 1440

QY 3887----- 3886

Db 1441 LQCEL SKAAPVWRKGSSETLRDGDYSRLRDQGTCELCI RGLAVEDTGEYLCVCGQERTS 1500



```

QY 3887 ----- 3886
Db 1501 ATLTVRALPARFIDNMTNQAREGATATLHCELSKVAPVEMKGPETLRDGRHSLRQDG 1560
QY 3887 ----- 3886
Db 1561 TRCEQLIRGLSVADAGEYSCVCGQERTSATLTIREATGATAMQLCELSKVAPVEMKGP 1620
QY 3887 ----- 3897
Db 1621 ETLRDGRYLRQDTRCCEQLIHGLSVADTGEYSCVCGQERTSATLTIRKAPQVFRPELQ 1680
QY 3898 SLOABEGSTATLQCELSPTATVVMKGLQLOANGREPRPLQCTAELVLQDLQREDTG 3957
Db 1681 SLOABEGSTATLQCELSPTATVVMKGLQLOANGREPRPLQCTAELVLQDLQREDTG 1740
QY 3958 EYTCGSGQATSATLTVAAPVRLREI-OHQEVDEGGTAHLCCELSPAGASVEMKGSIQ 4017
Db 1741 EYTCGSGQATSATLTVAAPVRLREI-OHQEVDEGGTAHLCCELSPAGASVEMKGSIQ 1800
QY 4018 LFPCKAYQWQDGAALVVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFTRQLQSL 4077
Db 1801 LFPCKAYQWQDGAALVVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKFTRQLQSL 1860
QY 4078 EQETGDIARLCCQLSDAESGAVOWLKEGVELHAGPKYEMRSQATRELLIHOLEAKDTG 4137
Db 1861 EQETGDIARLCCQLSDAESGAVOWLKEGVELHAGPKYEMRSQATRELLIHOLEAKDTG 1920
QY 4138 EYACVTGQKTAASLRVT-EPEVTVIRGLVDAEVTADVEDFECEVSAGATGVQWCLQG 4196
Db 1921 EYACVTGQKTAASLRVTGEPEVTVIRGLVDAEVTADVEDFECEVSAGATGVQWCLQG 1980
QY 4197 LPLOSNEVTEVAVRDGRHITLRLKGVTPEDAGTVSFHLGNHASAQLTVRAPEVTILEPL 4256
Db 1981 LPLOSNEVTEVAVRDGRHITLRLKGVTPEDAGTVSFHLGNHASAQLTVRAPEVTILEPL 2040
QY 4257 QDVQLSEGQDASFCLRSRAGQARWALGVPLQANEMNDITVEQGTLLHLLHKVTLE 4316
Db 2041 QDVQLSEGQDASFCLRSRAGQARWALGVPLQANEMNDITVEQGTLLHLLHKVTLE 2100
QY 4317 DAGTVSHVGTCSSEAOIKVTAKTVVIRGLVDAEVTADVEDFECEVSAGATGVQWCLQG 4376
Db 2101 DAGTVSHVGTCSSEAOIKVTAKTVVIRGLVDAEVTADVEDFECEVSAGATGVQWCLQG 2160
QY 4377 DEPVRTSENAEVFFENGLRHLLKLNLRPQDSCRVTFLAGDMVTSAFLTVR 4428
Db 2161 DEPVRTSENAEVFFENGLRHLLKLNLRPQDSCRVTFLAGDMVTSAFLTVR 2212

RESULT 3
Q9HCD3
ID Q9HCD3 PRELIMINARY; PRT; 1319 AA.
AC O9HCD3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein KIAA1639 (Fragment).
GN KIAA1639.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirotsawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.;"
RL DNA Res. 7:273-281(2000).
DR EMBL; AB046859; BAB13465.1;
DR HSSP; P56276; ITLK.
GO; GO:0005524; F:ATP binding; IEA.

```

```

DR GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO:0016740; F:transferase activity; IEA.
DR GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003538; IG_c2.
DR InterPro; IPR000719; Prot Kinase.
DR InterPro; IPR008266; Tyr_kinase_AS.
DR Pfam; PF00047; ig_1.
DR Pfam; PF00069; pkinase; 2.
DR ProDom; PD000001; Prot_Kinase; 2.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00408; IGc2; 1.
DR PROSITE; PS00835; IG LIKE; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 2.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Hypothetical protein; ATP-binding; Immunoglobulin domain; Transferase.
FT NON_TER 1
SQ SEQUENCE 1319 AA; 142049 MW; FBA43AE17204EF48 CRC64;

Query Match 16.7%; Score 6902; DB 4; Length 1319;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1319; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 6648 AMGVTSYLSLTCCSPFAGESDRATLLNVLEGRVSWSPMAHLSEDAKDFIKATLQRAPO 6707
Db 1 AMGVTSYLSLTCCSPFAGESDRATLLNVLEGRVSWSPMAHLSEDAKDFIKATLQRAPO 60
QY 6708 ARPSAAOCLSHWPFLKSNPAAEAFINTKQLKFLARSRWORSIMYSKILVMRSIPELL 6767
Db 61 ARPSAAOCLSHWPFLKSNPAAEAFINTKQLKFLARSRWORSIMYSKILVMRSIPELL 120
QY 6768 RGPDPSPSLGVARHLCRDTGGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 6827
Db 121 RGPDPSPSLGVARHLCRDTGGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 180
QY 6828 PSASLPEAEASESESTEAPAPSPGEGAGPPAAOCVPRHSVIRSLFVHQAGSEPEHAL 6887
Db 181 PSASLPEAEASESESTEAPAPSPGEGAGPPAAOCVPRHSVIRSLFVHQAGSEPEHAL 240
QY 6888 APGRRRHPRARRHLKGGYIAGALPGLREPLMEHRVLEEEAAREEQATLLAKAPSFETAL 6947
Db 241 APGRRRHPRARRHLKGGYIAGALPGLREPLMEHRVLEEEAAREEQATLLAKAPSFETAL 300
QY 6948 RLPASGTHLAPGHSHSLHSDSPSTPRSSSEACGAQRLPSAPSGAGIIRDMGHQGSQOL 7007
Db 301 RLPASGTHLAPGHSHSLHSDSPSTPRSSSEACGAQRLPSAPSGAGIIRDMGHQGSQOL 360
QY 7008 PSTGCHPGTAQPERPSPDPSPMGQAPFCHPKQGSAPQEGGSPHPAVAPCPGPPGSCCK 7067
Db 361 PSTGCHPGTAQPERPSPDPSPMGQAPFCHPKQGSAPQEGGSPHPAVAPCPGPPGSCCK 420
QY 7068 EAPLVSPSPFLGQQAAPPAPAKASPLDLSKMGPGDISLPGRPKPGCPSSPGCSAQSSSQ 7127
Db 421 EAPLVSPSPFLGQQAAPPAPAKASPLDLSKMGPGDISLPGRPKPGCPSSPGCSAQSSSQ 480
QY 7128 VSSLRVGSSQVGTPTGPGSLDAEGWTOAEADLSDSTPTLQRPQEQATWKRKSLGGRGYAG 7187
Db 481 VSSLRVGSSQVGTPTGPGSLDAEGWTOAEADLSDSTPTLQRPQEQATWKRKSLGGRGYAG 540
QY 7188 VAGYGTFAFGDAGGMLQGQPMWARIAWVSQSEEEQEEARAEBSQSEEQEAAESPLP 7247
Db 541 VAGYGTFAFGDAGGMLQGQPMWARIAWVSQSEEEQEEARAEBSQSEEQEAAESPLP 600
QY 7248 QVSARPPVEVGRAPTRSSPEPTWEDIGQVSLVQIRLSDGDAEAADTISLSDISEVDPAYL 7307
Db 601 QVSARPPVEVGRAPTRSSPEPTWEDIGQVSLVQIRLSDGDAEAADTISLSDISEVDPAYL 660
QY 7308 NLSLDYDKLPPFFMIFRKYKPSAQPEPPSPMAEEELAEFFPTWPPGELGPHAGLEI 7367
Db 661 NLSLDYDKLPPFFMIFRKYKPSAQPEPPSPMAEEELAEFFPTWPPGELGPHAGLEI 720
QY 7368 TESESDVDALLAEAAVGRKKNSSPSRSLFHFGRHLPLDEPAELGLREVRKASVEHISR 7427

```



Db	721	TEESDVEDALLAEAAVGRKWKSPSSLSLFFHFGPHLPDPEAELGLRERVKASVEHIS	780
Qy	7428	ILKRPGLKEGPRKKPGLASRLSGLKSWDAPTFIRELSDETIVLGOSVTLACQVS	7487
Db	781	ILKRPGLKEGPRKKPGLASRLSGLKSWDAPTFIRELSDETIVLGOSVTLACQVS	840
Qy	7488	AQPAQAQATWSKDGAFLSSSSRLVLSATLKNFQLTILVVAEDLGVYTCVSNALGTVT	7547
Db	841	AQPAQAQATWSKDGAFLSSSSRLVLSATLKNFQLTILVVAEDLGVYTCVSNALGTVT	900
Qy	7548	TGVLKKAERPSPPCPDGEIVADGVLLVWKPVESYGPVYIVQCSLEGSSWTLASDIF	7607
Db	901	TGVLKKAERPSPPCPDGEIVADGVLLVWKPVESYGPVYIVQCSLEGSSWTLASDIF	960
Qy	7608	DCCVLTSLKSRGGTYTERTACVSKAGMGPYSPSEQVLLGSPSHLASEESQGRSAQPLP	7667
Db	961	DCCVLTSLKSRGGTYTERTACVSKAGMGPYSPSEQVLLGSPSHLASEESQGRSAQPLP	1020
Qy	7668	STKTFAPQTQIRGRFSVVRQCEKASGRALAAKIIIPYHPKDKTAVLREVEALKGLRHPH	7727
Db	1021	STKTFAPQTQIRGRFSVVRQCEKASGRALAAKIIIPYHPKDKTAVLREVEALKGLRHPH	1080
Qy	7728	LAQLHAAVLSRHLVLIILELCSGPELLPCLAEASYSSESEVKDYLMQWLSATQYLHNOHI	7787
Db	1081	LAQLHAAVLSRHLVLIILELCSGPELLPCLAEASYSSESEVKDYLMQWLSATQYLHNOHI	1140
Qy	7788	LHLDLRSENMIITEYNLLKVVLDLGNQASLSQEKVLPSPDKXDKYLETWAPELLEQCGAVPQ	7847
Db	1141	LHLDLRSENMIITEYNLLKVVLDLGNQASLSQEKVLPSPDKXDKYLETWAPELLEQCGAVPQ	1198
Qy	7848	TDIWAIGVTAFTIMLSAEYFVSSEGARDLQRLKGLVRLSCYAGLSGGAVAFRLSTLCA	7907
Db	1199	TDIWAIGVTAFTIMLSAEYFVSSEGARDLQRLKGLVRLSCYAGLSGGAVAFRLSTLCA	1258
Qy	7908	QPMWGPCASSLCCQPLWTEEGPACSRPAPVTFPTARLRFVVRNREKRALLYKXHNLAQV	7967
Db	1259	QPMWGPCASSLCCQPLWTEEGPACSRPAPVTFPTARLRFVVRNREKRALLYKXHNLAQV	1318
Qy	7968	R 7968	
Db	1319	R 1319	

RESULT 4

Q9HCL6

Q9HCL6 PRELIMINARY; PRT; 1596 AA.

AC Q9HCL6; TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Hypothetical protein KIAA1556 (Fragment).

GN KIAA1556.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

SEQUENCE FROM N.A.

RP MEDLINE=20450683; PubMed=10997877;

RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes.

RT XVIII. The complete sequences of 100 new cDNA clones from brain which

RT code for large proteins in vitro."

RL DNA Res. 7:273-281(2000).

DR EMBL; AB046776; BAB1382.1; -.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003598; Ig\_c2.

DR Pfam; PF00047; Ig; 17.

DR SMART; SM00408; IgC2; 7.

DR PROSITE; PS50835; IG\_LIKE; 15.

KW Hypothetical protein; Immunoglobulin domain.

FT NON\_TER 1

FT	NON TER	1596	1596	
SQ	SEQUENCE	1596 AA;	175280 MW;	89937E0760C4E2EF CRC64;
	Query Match	16.0%;	Score 6619.5;	DB 4; Length 1596;
	Best Local Similarity	81.7%;	Pred. No. 0;	
	Matches 1309;	Conservative 70;	Mismatches 203;	Indels 21; Gaps 7;
Qy	2726	RARVRVHDLHVGITKRLKTMVELEGESCSFEVLSHESASDPAMTVGKTVGSSRRFOA	2785	
Db	1	RARVRVHDLHVGITKRLKTMVELEGESCSFEVLSHESASDPAMTVGKTVGSSRRFOA	60	
Qy	2786	TQGRKYIILVREAAAPSADAGEVVFVSRGLTSKASLIIVREPAALIKPLEDQWVAFGEVDE	2845	
Db	61	TQGRKYIILVREAAAPSADAGEVVFVSRGLTSKASLIIVREPAALIKPLEDQWVAFGEVDE	120	
Qy	2846	LRCELSRAGTPVHMLKDKRAIKRSQKYDVVCEGTWMLVIRGASLKADAGEYTCVEASKS	2905	
Db	121	LRCELSRAGTPVHMLKDKRAIKRSQKYDVVCEGTWMLVIRGASLKADAGEYTCVEASKS	180	
Qy	2906	TASLHVEEKANCFTBELTNLQVEEKGTAVFTCKTEHPAATVTRKGLLELRASGRKHQPSQ	2965	
Db	181	TASLHVEEKANCFTBELTNLQVEEKGTAVFTCKTEHPAATVTRKGLLELRASGRKHQPSQ	240	
Qy	2966	EGLTLRLTTSALKADSDTYTCDIGQAQSRALLVOGRVHIIEDLEDVDVQEGSSATFR	3025	
Db	241	EGLTLRLTTSALKADSDTYTCDIGQAQSRALLVOGRVHIIEDLEDVDVQEGSSATFR	300	
Qy	3026	CRISPANYPVHFWLTKTPHANELNEIDAQPGYHVLTRQALAKDSCTIYFEAGDQRA	3085	
Db	301	CRISPANYPVHFWLTKTPHANELNEIDAQPGYHVLTRQALAKDSCTIYFEAGDQRA	360	
Qy	3086	SAALRYTEKPSVFSRRLTDTATITEGSDTLVCTETCDIPMCWTGDKTLRGSARCOLSH	3145	
Db	361	SAALRYTEKPSVFSRRLTDTATITEGSDTLVCTETCDIPMCWTGDKTLRGSARCOLSH	420	
Qy	3146	EGHRAQLLITGATLODSGRYKCEAGACSSSIVRVHARVPRFOEALKDLEVEGGAAATLR	3205	
Db	421	EGHRAQLLITGATLODSGRYKCEAGACSSSIVRVHARVPRFOEALKDLEVEGGAAATLR	480	
Qy	3206	CVLSSVAAPVKWCYGNVLRPGDKYSLROEGAMLELVVNRNLRPDQSGRYSCSFGDQTTSA	3265	
Db	481	CVLSSVAAPVKWCYGNVLRPGDKYSLROEGAMLELVVNRNLRPDQSGRYSCSFGDQTTSA	540	
Qy	3266	TLVTALPAQFIGIKLRNKEATGATATLRCELSTAPVWRKSGSETLRDGDYCLRDQGA	3325	
Db	541	TLVTALPAQFIGIKLRNKEATGATATLRCELSTAPVWRKSGSETLRDGDYCLRDQGA	600	
Qy	3326	MCELOIRGLAMVDAAEYSCVCGEERTSASLTIRPMPAHFIRGRHOSIEGATATLRCEL	3385	
Db	601	MCELOIRGLAMVDAAEYSCVCGEERTSASLTIRPMPAHFIRGRHOSIEGATATLRCEL	660	
Qy	3386	SKAAPVWRKGRSRLDRGDRHSLRQDGA VCELOICGLAVADAGEYSCVCGEERTSATLT	3445	
Db	661	SKAAPVWRKGRSRLDRGDRHSLRQDGA VCELOICGLAVADAGEYSCVCGEERTSATLT	720	
Qy	3446	KALPAKFTGRLNEEAVEGATAMLWCELSKVAPVWRKSGPENLRDGDYTLRQEGTRCEL	3505	
Db	721	KALPAKFTGRLNEEAVEGATAMLWCELSKVAPVWRKSGPENLRDGDYTLRQEGTRCEL	780	
Qy	3506	QICGLAVADAGEYLCVCGQERTSATLTIRALPARFIEDVKNQAREGATAVLOCELNSAA	3565	
Db	781	QICGLAVADAGEYLCVCGQERTSATLTIRALPARFIEDVKNQAREGATAVLOCELNSAA	840	
Qy	3566	PVWRKSGSETLRDGDYSLRQDGTCKELQIRGLAMADTGEYSCVCGQERTSAMLTVRALP	3625	
Db	841	PVWRKSGSETLRDGDYSLRQDGTCKELQIRGLAMADTGEYSCVCGQERTSAMLTVRALP	900	
Qy	3626	IKFTGRLNEEAVEGATAVLRCELKSNAPVWRKSGHETLDRGRHSLRQDGA RCELQIRG	3685	
Db	901	IKFTGRLNEEAVEGATAVLRCELKSNAPVWRKSGHETLDRGRHSLRQDGA RCELQIRG	960	
Qy	3686	LVAEDAGEYLCVCGKERTSAMLTVRAMPKFIIEGLRNEEATGDTATLWCELSKAAPVW	3745	

Db 961 LVAEDAGEYLCWCGKERTSAMLTVRAMPKSFIEGLRNEATEGDTATLWCBLSKAAAPVEW 1020  
 QY 3746 RKGHETLRDGRHSRLQDGRSCELQIRGLAVVDAGEYSCVCGQERTSATLTVRALPARFI 3805  
 Db 1021 RKGHETLRDGRHSRLQDGRSCELQIRGLAVVDAGEYSCVCGQERTSATLTVRALPARFI 1080  
 QY 3806 EDVKQAEAREGATVLOCELKAAAEVWRKSGSETLRGGDRYSLRQDGRTRCQLQIHGSLVA 3865  
 Db 1081 EDVKQAEAREGATVLOCELKAAAEVWRKSGSETLRGGDRYSLRQDGRTRCQLQIHGSLVA 1140  
 QY 3866 DTGEYSCVCGQERTSATLTVRAPOVFRPEPLQSLQAEBSGSTATLOCELSEPTATVWVSKG 3925  
 Db 1141 DTGEYSCVCGQERTSATLTVRALPARFQDLTKAESEGATATLOCELK--VAPVWRKG 1199  
 QY 3926 GLQLANGREPRLOGCTAELVLODLORED7GEYTCGSGQATSATLTVTAAPVRFLEL 3985  
 Db 1200 PETLRDGGYSLKQDGRTRCQLQIHLSVADAGEYSCMCGQERTSATLTVRALPARFTEGL 1259  
 QY 3986 OHQEVDEGTAHLCCELSPAGASVWRKSGSLQLPCKAYQMVQDGAALVVRGVEQEDA 4045  
 Db 1260 RNEEAMEGATVLOCELKSA--APVWRKGLSALRDGDKYSLRQDGAVALQIHGLAMADN 1318  
 QY 4046 GDYTCDTGHTOSMASLSYVRPRPKTRLOSLEQETGDIARLCCOLSDASSGAVVOMLKE 4105  
 Db 1319 GYVSCVCGQERTSATLTVRALPARFIEDMRNKGATGATVTLQCKLRKA---APVWRKG 1375  
 QY 4106 GVHLLHNGPKYMRSGATRELLIHQLEAKDTGEVACTGCGKTAASLRVTEPEVTIVRGL 4165  
 Db 1376 PNTLKQDGRYSLKQDGRTRCQLQIRGLVIADAGEYSCICEQERTSATLTVRALPARFIEDV 1435  
 QY 4166 VDAEYTABEDVEFSCVSRAAGATGVQWCLQGLPLQSNTEVAVRDGRIHTLRKLGVTPE 4225  
 Db 1436 RNHEATEGATVLOCELKSAAP--VWRKSGSETLRDGRYSIR--QDTRCQLQIRGLAVE 1492  
 QY 4226 DAGTVSFHLGNHSAQITVRAPEVTILEPLQDVLQSEGQDASPQCRSLRASQAEARWAL 4285  
 Db 1493 DTGEYLCVCGQERTSATLTVRALPARFIDNMNQAEREGATATLHCELSKVAPVE--WRK 1550  
 QY 4286 GGVLQANEMNDITVEQGTLLHLLHKVTLLEDAGTVSFHVGTC 4328  
 Db 1551 GPEILRDGRHSRLQEN-----RLNPGGGCGSELGSC 1582

## RESULT 5

Q8NH2  
 ID Q8NH2 PRELIMINARY; PRT; 1040 AA.  
 AC Q8NH2;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Obscurin (fragment).  
 GN OBSN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Young P., Ehler E., Gautel M.;  
 RT "Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere assembly";  
 RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ314506; CAC85753.1; --  
 DR InterPro; IPR003962; FnIII\_subd.  
 DR InterPro; IPR003961; FN III.  
 DR InterPro; IPR008957; FN III-like.  
 DR InterPro; IPR003599; Ig.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig-c2.  
 DR InterPro; IPR000048; IQ-region.  
 DR Pfam; PF00041; fn3; 1.  
 DR Pfam; PF00047; ig; 6.  
 DR Pfam; PF00612; IQ; 1.

DR PRINTS; PR00014; FNTYPEIIII.  
 DR SMART; SMC00060; FN3; 1.  
 DR SMART; SMC00409; IG; 6.  
 DR SMART; SMC00408; IGC2; 5.  
 DR SMART; SMC00015; IQ; 1.  
 DR PROSITE; PS00835; IG\_LIKE; 5.  
 DR PROSITE; PS00096; IQ; 1.  
 KW Immunoglobulin domain; Repeat.  
 FT NON\_TER 1  
 FT TER 1040 1040  
 SQ SEQUENCE 1040 AA; 112354 MW; 3DE879CA266D7F44 CRC64;

Query Match 13.0%; Score 5358.5; DB 4; Length 1040;  
 Best Local Similarity 99.4%; Pred. No. 2.6e-279;  
 Matches 1034; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 4430 WRLEILPLKNAAVRAGAAQARTCTLSEAVPVGEASWYINGAAVQPDSDMTVTADGSHQ 4489  
 Db 1 WRLEILPLKNAAVRAGAAQARTCTLSEAVPVGEASWYINGAAVQPDSDMTVTADGSHH 60  
 QY 4490 ALLLSAQPHHAGEVTFACRDAVASARLTVLGLPDPPEDAEVVAHSSHTVLSWAAPMSD 4549  
 Db 61 ALLLSAQPHHAGEVTFACRDAVASARLTVLGLPDPPEDAEVVAHSSHTVLSWAAPMSD 120  
 QY 4550 GGGGLCGYRVEKEGATQWRILCHELVGPPCVDGLAPGETYPRVAAGVPGAGEPVH 4609  
 Db 121 GGGGLCGYRVEKEGATQWRILCHELVGPPCVDGLAPGETYPRVAAGVPGAGEPVH 180  
 QY 4610 LPQTVRLAEPKPPVPPQSPAPESROVAAGEDVSLSEVVAAGVIVHKGHERIQPGGRF 4669  
 Db 181 LPQTVRLAEPKPPVPPQSPAPESROVAAGEDVSLSEVVAAGVIVHKGHERIQPGGRF 240  
 QY 4670 FVVSQGRQMLVIKFTAEADQGEYHGLAQSICPAATFOVALSPASVDEAPQSLPPE 4729  
 Db 241 FVVSQGRQMLVIKFTAEADQGEYHGLAQSICPAATFOVALSPASVDEAPQSLPPE 300  
 QY 4730 AAQEGDLHLLEALARKMRSEPTLDSISLPEDGSRQRLPOEAEVAPDLSEGYSTA 4789  
 Db 301 AAQEGDLHLLEALARKMRSEPTLDSISLPEDGSRQRLPOEAEVAPDLSEGYSTA 360  
 QY 4790 DELARTGDADLSHTSSDDESRAAGTSLVTLKAGRPCTSLASKVGAAPSVKPPQQQ 4849  
 Db 361 DELARTGDADLSHTSSDDESRAAGTSLVTLKAGRPCTSLASKVGAAPSVKPPQQQ 420  
 QY 4850 EPLAAVRPPLGDLSTKDLGDPMDKAAVKIQAAKGYKVRKEMKQOQSGPFSHTFGTEA 4909  
 Db 421 EPLAAVRPPLGDLSTKDLGDPMDKAAVKIQAAKGYKVRKEMKQOQSGPFSHTFGTEA 480  
 QY 4910 QVGDALRLCECVVASKADVRARWLKDGVELTDGRHHIDQLDGDGTCSLLIAGLDRADAGCY 4969  
 Db 481 QVGDALRLCECVVASKADVRARWLKDGVELTDGRHHIDQLDGDGTCSLLIAGLDRADAGCY 540  
 QY 4970 TCQVSNKFGQVTHSACVVSSESESSSGGELDDAFRAARLHLRLFRTKSPAESVDE 5029  
 Db 541 TCQVSNKFGQVTHSACVVSSESESSSGGELDDAFRAARLHLRLFRTKSPAESVDE 600  
 QY 5030 ELFLSADGPAEPPEPADWQTYREDEHFCIRFPALEARQAVTRFOEMFATLIGIVEIK 5089  
 Db 601 ELFLSADGPAEPPEPADWQTYREDEHFCIRFPALEARQAVTRFOEMFATLIGIVEIK 660  
 QY 5090 LVEQGRPRVEMCISKETAPVVPPEPLSLTSDAAPVFLTELQNEQVQDGPVSPFCVV 5149  
 Db 661 LVEQGRPRVEMCISKETAPVVPPEPLSLTSDAAPVFLTELQNEQVQDGPVSPFCVV 720  
 QY 5150 TGQPMPSVRWFKDGLLEDDHYMINEDQGGHQLIITAVVPADMGVYRCLAENMGVSS 5209  
 Db 721 TGQPMPSVRWFKDGLLEDDHYMINEDQGGHQLIITAVVPADMGVYRCLAENMGVSS 780  
 QY 5210 TKAEALVDLTSTDYDTAADATESSSYSFSAQGYLS--SREQEGTSTTDEGQLPQWVEELRD 5268  
 Db 781 TKAEALVDLTSTDYDTAADATESSSYSFSAQGYLS--SREQEGTSTTDEGQLPQWVEELRD 840  
 QY 5269 LQVAPGRTLAKFQLKVKGYPAAPRLYWFKQOPLTASAHIRMTGCKILHTLLEIISVTRDS 5328

```
Db 841 LQVAFGTFLAKFQKVGYPAPRYWFKOGQFLTSASAHRTMDKKILHTLEIISVTRDS 900
QY 5329 GQYAAIYNAMGAAYSSARLLVRCGPDEPEEPFASDVHQLVPPRMLERFPPKVKKGSSI 5388
Db 901 GQYAAIYNAMGAAYSSARLLVRCGPDEPEEPFASDVHQLVPPRMLERFPPKVKKGSSI 960
QY 5389 TFSVKVEGRPPTVHVLREAEERGVNLWIGPDPGYTVASSAQOHSVLVLDVGRHQGTYT 5448
Db 961 TFSVKVEGRPPTVHVLREAEERGVNLWIGPDPGYTVASSAQOHSVLVLDVGRHQGTYT 1020
QY 5449 CIASNAAGQALCSASLHVS 5468
Db 1021 CIASNAAGQALCSASLHVS 1040

RESULT 6
Q86YC7 PRELIMINARY; PRT: 767 AA.
AC Q86YC7;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Similar to titin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Straussberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC043916; RAH43916.1; -.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; -IG-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00047; IG; 5.
DR SMART; SM00409; IG; 7.
DR SMART; SM00408; IGC2; 5.
DR PROSITE; PS50835; IG_LIKE; 5.
FT NON_TER
SQ SEQUENCE 767 AA; 84732 MW; 39AB316E63964DAA CRC64;

Query Match 9.4%; Score 3865; DB 4; Length 767;
Best Local Similarity 99.5%; Pred. No. 3.7e-199;
Matches 751; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2248 VKVQGRTYTLIYRRVLAEDAGBIQFVAENASRAQLRVKELPVTILVRPLRDKIAMEXHRG 2307
Db 5 IRQEGRTYTLIYRRVLAEDAGBIQFVAENASRAQLRVKELPVTILVRPLRDKIAMEXHRG 64
QY 2308 VLECOVSASAOVRFKSGOELQPGKYELVSDGLYRKLIIISDVHAEDEDTYTCADGVK 2367
Db 65 VLECOVSASAOVRFKSGOELQPGKYELVSDGLYRKLIIISDVHAEDEDTYTCADGVK 124
QY 2368 TSAQFVEEQSTIVRGLODVTWMPAPAWPFECETSIPIVPPKWLKTKTVLQAGNVGL 2427
Db 125 TSAQFVEEQSTIVRGLODVTWMPAPAWPFECETSIPIVPPKWLKTKTVLQAGNVGL 184
QY 2428 EOGGVTHRLMRLRTCTMTGPVHFTVKGSRSSARLVSDIPVLTFRPLEPKTGRELSQV 2487
Db 185 EOGGVTHRLMRLRTCTMTGPVHFTVKGSRSSARLVSDIPVLTFRPLEPKTGRELSQV 244
QY 2488 LSCDFRPAKQVQWYKDDTPLSPSEKFKVMSLEGQVAELRIILMPADAGVYRCQAGSAHS 2547
Db 245 LSCDFRPAKQVQWYKDDTPLSPSEKFKVMSLEGQVAELRIILMPADAGVYRCQAGSAHS 304
QY 2548 STEVTVEAREVTVTGLQDAEATERGWASFSCLSHEDEEVEWSLNGMPLYNDSPEISH 2607
Db 305 STEVTVEAREVTVTGLQDAEATERGWASFSCLSHEDEEVEWSLNGMPLYNDSPEISH 364
QY 2608 KGRHRTLVLKSIORADAGIVRASSLVKSTARSLEVRVPVFLKALDDLSAEERGTALQ 2667
```

## RESULT 7

```
Q8WZB3 PRELIMINARY; PRT: 26926 AA.
ID Q8WZB3;
AC Q8WZB3;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE N2B-titin isoform.
GN TTN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20309627; PubMed=10850961;
RA Freiburg A., Trombitas K., Hell W., Cazorla O., Fougereousse F.,
RA Centner T., Kolmerer B., Witt C., Beckmann J.S., Gregorio C.C.,
RA Granzier H., Labeit S.;
RT "Series of exon-skipping events in the elastic spring region of titin
RT as the structural basis for myofibrillar elastic diversity.";
RL Circ. Res. 86:1114-1121(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=21573839; PubMed=11717165;
RA Bang M.L., Centner T., Fornoff F., Geach A.J., Gotthardt M.,
RA McNabb M., Witt C.C., Labeit D., Gregorio C.C., Granzier H.,
RA Labeit S.;
RT "The complete gene sequence of titin, expression of an unusual -700
RT kDa titin isoform and its interaction with obscurin identify a novel
RT Z-line to I-band linking system.";
RL Circ. Res. 89:1065-1072(2001).
CC -1- SIMILARITY: CONTAINS 91 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
DR EMBL; AJ277892; CAD12455.1; -.
DR GO; GO:0016020; C.membrane; IEA.
DR GO; GO:0005224; F.Arp binding; IEA.
DR GO; GO:0004896; F.hematopoietin/interferon-class (D200-domain. . .; IEA.
DR GO; GO:0004601; F.peroxidase activity; IEA.
DR GO; GO:0004674; F.protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; F.protein-tyrosine kinase activity; IEA.
DR GO; GO:0016740; F.transferase activity; IEA.
DR GO; GO:0005975; P.carbohydrate metabolism; IEA.
DR GO; GO:0006468; P.protein amino acid phosphorylation; IEA.
DR GO; GO:0006979; P.response to oxidative stress; IEA.
DR InterPro; IPR000282; Cytok_receptor_2.
```

Db	1625	EIAAPELEPLHLRYQBQWEEGLDYKEKQKQPFKKKLTSLRKFGPAHFRCRLTPIG	1684
Qy	358	QPSTEAAWFKEETRL-----WASAKYGI-----	EEEGTERRLT 390
Db	1685	DPTMVVSWLHDGKPLEANLRMINFEGYCSLDYGVAYSRDSGIIICRATNKYTGTDHTSA	1744
Qy	391	VRNVSADDDAVYICETPEGSR-----TVAELAVQGNL-----	LRKLPRKXT 430
Db	1745	TLIVKDEKSLVEESQLPEGRKGLQRIEELERMAHEGALTCVTTDQKEKQKPDVLYPEPV	1804
Qy	431	AVRVGDTAMP-CVELAVPGPVHWRNQBEVAVAGGRVAISABGTRHTLTIISQCCLDEVQ	489
Db	1805	RVLEGETARFCRCRVTPGQPKVNWYINGQLIRKSKFRVRYDGI-HYLDIVDCSKSYDTGE	1863
Qy	490	VAFMAG-----DCQTSRFCVSAPRKPPLQP-----	PVD 518
Db	1864	VKVTAEENPEGVIEHKVKLEJQQREDFRSLRRAPRPRFHVHPEKQLOFEVKQVDRPVD	1923
Qy	519	-----PVVKARMSSVILSWSPPHGE-----RPVTIDGY-----	LVEKKLGTYYW 560
Db	1924	TTETKEVVKLRAERITHEKVPESBELRSKFRRTTEEGYEAITAVELKSRKODESYEE	1983
Qy	561	I-----RCHAEAWATPELTIVAD--VAEENQFRVSALNSFCQSPYLEFPFGTVHLPK	613
Db	1984	LLRKTKDELLHW--TKELTEEEKKALAEGBKITPTFKPKDKIELSMS-----	APKI 2034
Qy	614	AVRPLKAVQAVEGGEVTFSDVLTVASAG-----BWFLDGOALKASS-VYEHCDRTSH	666
Db	2035	FERIQSQTVG--QCSDAHFRVR-----VGKPDPECEWYKNGVKIERSDRIYVWPDNVC	2088
Qy	667	TLTIREVPASLHGAQKLF--VANGIESIEMEVRAAPGLTANKPPAPAAAREVLARLHE	722
Db	2089	ELVIRDVTAE-DSASTWVKAINAGTSSHAFLVQAKOLITFTQ-----ELQDVAKED	2143
Qy	723	E-AQLIAELSDQAAAVTMLKDQRTLSPGPKYEQVASAGRRVLLRDVARDADGLYECV--	779
Db	2144	TMATFECETSEPFVKVKYKQMGVEHGDYKRMHSDRKVHFLSILTIDTSDAEDYSCVLV	2203
Qy	780	-SRGGRIANQLSVOGLARFLHKDMAGSCVDVAVAGGPAQPECETSEAHVHVHVHVKDMELG	838
Db	2204	EDENVKTKAKLIVEGAUVFVKELQD--TEVPESYSGELECIUSPENIEGKWHYNDVELX	2261
Qy	839	HSGERFLQEDVGTRHRLVAATVTRQDEGYTSRCRVGEDSDVDFRLRVSEPKVPFAKQLARR	898
Db	2262	SNCKYIITSRRG-RQNLTKVDTKEDQGEYSFVIDGKTTCKLKW-KRPPIALQGLSDQ	2319
Qy	899	KLOAEAGASATLSCFAVAQAQTEVTWYKDGKLSKSSKVCNEATGTRRLRVQOQAQAD	958
Db	2320	KVC--BGDIVQLEVKVLSLESEGVKMWKQGOQPSDRVHIVIDKQSHMLLIIDMTKED	2377
Qy	959	EYS-----CBAGGORLSFHLIDKPEKVPVFAKDQVAHSEVOAEAGANATLSCFAQAQ-	1010
Db	2378	NYSFPTPALGLSTSGRVSYVDVITPL-----KDVNVIETRAVLECKYSVPDV	2427
Qy	1011	AEVWYKDGKLSKSLUKVHVEAKGRRRLRVQOAKTDAGDYSCEARGORVSRFRIHTEP	1070
Db	2428	TSVKWYINDEQIKPDRVOAIKVGTKQRLVIRNTHASDEGPYKLIIVGRVETNCNLSVEKI	2487
Qy	1071	KXWFAKEQSVHNEVOAEAGASAMLSCEVAQAQTEVTWYKDGKLSKSSKVGMEVGCTRR	1130
Db	2488	KIIRGRLDITCTETQ-----NVVFEVLSHSGIDVLWFPKDEIKPSSKYKEANGKIYK	2542
Qy	1131	LVLQPAGKADAGEYSCBAGQORVSRFHLHTEPKGVFAKEQSVHNEVOABAGGTAMLSC	1190
Db	2543	LTVLNMKMDGKTYFYAGENITTSGLK--TVAGGAISK--PLTDQTVAES-CEAVFECEV	2597
Qy	1191	AQPQTEVTWYKDGKLSKSSKVRMEVKGCTRELVLVQOVGKADAGAGEYSCBAGQORV	1250
Db	2598	ANPDSGEMWRDGKHLPLTNIRSSDGHKRLIIAATKLLDIGBYTVKATSKTSAKLK	2657
Qy	1251	ITEPKVAFKQELVHNEVREAGASATLSCFAQAQTE-VTWYKDGKLSKSSKVRIBAA	1309

Db 2658 V---BAVKIKKTL-KNLTVTBT-QDAVFTVLTBPNVKGWQWIKNGVVLNENKVAISVK 2712  
 Qy 1310 GCMRLVVQQAQADAGEYTCAG--GQRLSHLD---VSEPKAVFAKEQLAHRKVOAE 1363  
 Db 2713 GTIYSLRINKAIVDES VYGRFLRGASARLHVETVKI IKKPKDVTALEN----- 2763  
 Qy 1364 AGATATLSCVAQAQ--TEVTWYKDGKLLSSSKVRMEAVGCTRLVVQQAQADTCFYS 1421  
 Db 2764 ---ATVAFEVSVSHDTPVFKWFKHKNVEIKPSDKHRLVSEKVKHKLMLQNLSPSDAGEYT 2819  
 Qy 1422 CEAGGQRLSFLDVAEPKVPFAKQPVH-----REVQAQAGASTLSCVAQAQATEVWVY 1476  
 Db 2820 AVVGQLECKAKLFV-----ETLHITKTMKNIEVETKTASFECEVSHFVPSMWL 2869  
 Qy 1477 KDGKLLSPSKVRMEAVGCTRLVVQQAQADAGEYSCAQSQRSLSHLHVABKAVFAK 1536  
 Db 2870 KNGVEIEMSEFKI VVQKQLHLIIMNTSTEDSAEYTFVCGNDQVSATLTVT-BIMITS- 2927  
 Qy 1537 EOPASREVQAAGT SATLSCVAQAQATEVTWYKDGKLLSSSKVRMEAVGCTRLVVQEA 1596  
 Db 2928 ---MLKDNAEKKUTITEVTWYEGISYKWLKNGVEIKSTDKQMRTKLTHSLNRNV 2984  
 Qy 1597 GOADAGEYSCAGDORLSFHLHVABPKVFAKEQPAHREVQAEAGASATLSCVAQAQTE 1656  
 Db 2985 HFGDAADYTFVAGKATSTATLYVEARHIEFRKH---INKIKVLEKGRAMFECEVSEPDIT 3041  
 Qy 1657 VTWYKDGKLLSSSKVRMEAVGCTRLVVQQAQADAGEYSCAQSQRSLSHLHVABE 1716  
 Db 3042 VQWMDDELQITDRIKQKQYVHRLIIPSTRMSDAGKITVAVAGNVSTAKLVFGRDV 3101  
 Qy 1717 QISERPCRRPLVKEHEDIILTATLTPSAATVTLKDGVEI---RRSKRHETASQGDTH 1774  
 Db 3102 RI--RSIKKEVQVIEKQAVV--BEVNEDDVAHWYKDGIEINFQVERKHYVVERRIH 3157  
 Qy 1775 TLTVHGAQVLSAISYCRVAGQDPVQVBE-EVAAKFCRLLEBVCGLGTVTVLACELS 1833  
 Db 3158 RMFISETRQSDAGEYTFVAGNRSSVTLVYNAPPEPQVLQELPQVTVQSG-----X 3208  
 Qy 1834 PA--CAEVVWRCGNTQPRV-----GKRFQWAEVGRVSRILTVLGLRAEDAGBYVC 1880  
 Db 3209 PARFCAVI---SGRPQPKISWYKEBQLLSTGFKCKFLHDGQEYTLILLIEARPEDAAYTC 3265  
 Qy 1881 ESRDHP-----TSAQITVSVPRVVK-----FMGLSTVVAEGGEATFCQCVSPS 1925  
 Db 3266 EAKNDYGVATTASLSVPEVSPQEMPVYPAITFLQDVTVTSEGPAPFCRVSGT 3325  
 Qy 1926 DVAVVWFRDGLALQSEKFAISQSGASHLATSILVLEDAGQITVEABGA---SSSAAL 1981  
 Db 3326 DLKVSWSKDKKIKSRFRMTQFEDTYQLBIAEAYPEDEGTYTFVASNAGVQVSSANL 3385  
 Qy 1982 VYREAPVLFKKLEPQ-----TVBERSSTVLELTER----- 2013  
 Db 3386 SLEAPESILHERIEQIEIEMEMKEFSSFLSAEEGLHSAELQLSKINETLELLSESPVYS 3445  
 Qy 2014 -----PWPELRWTRNATALAPCKNVEI 2035  
 Db 3446 TKFDESEKGTGPIFTKEVSNADISMGDVATLSTVTVIGIPKPKIOMFFNGVLLTBSADYKF 3505  
 Qy 2036 HAEGARHLVLNMGVADRGFGFCETPD----- 2066  
 Db 3506 VFDGDDHSLIILFTKLEDEGEYTCWASNDYKTCISAYLKINSKGEHGDHETESAVAKS 3565  
 Qy 2067 QAKL-----TVEMRQVRLVRLGQAV-EAREQGTATWEVOLSHADVGDGWTDRGLRFQ 2117  
 Db 3566 LEKLGCPCPPHFLKELKPIRCAQGLPAIFETVVGEPAPTV-----TWFKENKQL- 3615  
 Qy 2118 QGPTC---HLAVRGP--MHTLTLGLRPSDGLMVVFAEGV---HTSARLVV----- 2161  
 Db 3616 ----CTSVYVYTIHNPNGSGTIFVNDPQREDSGLYICAKENMLGESTCAEALLVLEDDTD 3671  
 Qy 2162 -----TELPVFSFR-PL-----QDVVTTTEKKEVLTCELSRPNVDVRW 2198  
 Db 3672 MTDTPCKAKSTPEAPEDFPQTLKGPVAVALDSEQEIATFVKDITILKAALITEENQOQLSY 3731

Qy 2199 -----LKQGVELPAGKTMATAAAG-----ACRSITTY----- 2225  
 Db 3732 EHIKANELSSQLPLGAQELQSLIODEKLTPTBSTREFLCINGSIHFPQLPKPSNQLQOI 3791  
 Qy 2226 ---RCFADQGVVVCDAHDAQS----- 2244  
 Db 3792 VQSQKTFSEKGLMPPEPETQAVLSTEKI FPSAMSIEQINSITVEPLKTLALAEPEGNY 3851  
 Qy 2245 SASVKVQGRTY-----TLIVRRVLAE----- 2265  
 Db 3852 QSSIEPPMHSYLTSAEVLSPKEKTVSDTNRQVRVTLQKBAQASALILQSILAEHVES 3911  
 Qy 2266 -----DAGEIQFVA---ENAESRAQLRVKE----- 2287  
 Db 3912 LQSPDVMISQVNVPELVPSEHSTEGKILISANPLENAGQDSAVRIEGLKSLRPLAL 3971  
 Qy 2288 -----LFTVLRPLRDXIAMKRGV----- 2308  
 Db 3972 EEQVLLKEEHSNVVMPDPQIIESKREPVAIKKQVQVRDLSKESLLSGIPEQRNL 4031  
 Qy 2309 LECQVSR- SAQV-----RW----- 2322  
 Db 4032 LKIQICRALQAAVASEQPLGFSEWLRNIEKVEAVNITQEPRHIMCMVLTSAKSVTBE 4091  
 Qy 2323 -----FKGSO 2327  
 Db 4092 VTIIIEDVDPQANLKWELRDALCAIYEEIDILTAEGPRIQOGAKTSIQEEMDSFSGQ 4151  
 Qy 2328 ELOGPKYELVSDGL-----YRKLISDVH-----AE 2354  
 Db 4152 KVBPITEPEVESKYLISPEEVSFVNSRVKYLDATPVTGASAVVSDEKQDESUKPBE 4211  
 Qy 2355 DEBTVTCDAG-----DVKTSQAPFVEBQSTIIVRGLODVTWMEPAPAFECETSI 2407  
 Db 4212 EKESSESSEGTETAVNVKIQEABGGIKEDGPMIHTPLVD-TVSEEGDI-VHLLTITNA 4269  
 Qy 2408 RPPKWLGLGTVLQAGNVGLQEGTVHRLMLRR-----TCSTMTGPVHVTGKSR 2457  
 Db 4270 KEVWYFENKLPSPDEKFKCLQDQNTYTLVIDKVNTEHQGEVCEALNDS-----GKTA 4324  
 Qy 2458 SSARL-VVSDI PVLTREPLEPTKRGELQSVLSCDPRPAPKA-VOMYKDDTPLSPSEKPK 2515  
 Db 4325 TSAKLTWVRAAPVIRKLEPLEVALGHLAKTCEIQSAPNVRFQFKAGRIEYSDKCS 4384  
 Qy 2516 MSLEGQMAELRILRLMPADAVYRCQAGSAHSESTVEAREVTVTG-----PLQD 2566  
 Db 4385 IRSKVISSLEILRLTQVDCGEYTCASNEYSVSCTA-TLTVTVPGGKKYKLLPERK 4443  
 Qy 2567 ABATEGNAFSCFELSHEDBEVWSLNGMPLINDSFHEISHKGRHRLVILKIQRADAGI 2626  
 Db 4444 PEPKEE--VVLKSVLAKRPEEBEPKVEPKL-----EKVKPAPVPEPPPKPVEVEVPT 4496  
 Qy 2627 VRASSLKVSTARSARLEVRVVPVFLKALDDLSABERGTLALQCEVSDPE-----AHV 2677  
 Db 4497 VTREKIKIPEPTKVP-EIKPAIPLPAPBPKPEAEVTKIKPPVEPETPIAPVTVFV 4555  
 Qy 2678 VWRKGVQGLSPSKYDFLHTAG-TRGLVVDVSPEDAGLYTCHVGS-----BETRARVR 2730  
 Db 4556 VGKKAQAK--APKEBAAPKGPPIKGVPKKTPSPIAEERKLRPGSGGKPPDEAPFTYQ 4612  
 Qy 2731 VHDLVHGTIKRLKTMETVLE-----GESCSPECVLSHESASDPAMWTVGKTVGSSSRFOAT 2786  
 Db 4613 LKAVPLFKVKEIKDIILITSEBFVGSSEIPECLVSPSTAI--TTWKMDGKNIESPKHRTI 4670  
 Qy 2787 ROGRKYLIVREAPSDAGEVWFSVR-----GLTSKASLIVRERPAIILKPLEDQ-WVAPG 2841  
 Db 4671 ADGKDRKLHIIDVQLSDAGEYTCVLRKGNKEKSTAKLVVEELVFRVFKTLEBEEVWVG 4730  
 Qy 2842 EDVELCELSRAGTFFHLLKORK-ATRKSKQVYVCEGTWAMVILVIGASLKDAGEYTCV 2900  
 Db 4731 QPLYLSCENLKE-RDVVWRKDGKIVVEKPRIVGVIGLMLRALTINDADTDAGTYTIV 4789



Db 6884 -----LGTLTADDAFVEPTMDLSAFKDGLEIVVPNPITILVPSTGY 6924  
Qy 4887 KVRKEMKQOEGPMESHFTGDTAEQVGDALRELCVUASKADVRAWLKDGVELTDCRHHI 4946  
Db 6925 -----PRETATWCFGDKVLETDGRVKMKTLSA----- 6951  
Qy 4947 DQJGDTCSLLIAGLDRADAGCYTCQVSN-----KFGQVTHS 4983  
Db 6952 -----YAEVLVSPSRSDKGIYTLKLENRVKTSIGEIDVNVVIARPSAPKELKFGDITKD 7005  
Qy 4984 ACVVVSGSESAESSGGELEDDAFRAARRLHURFRK----- 5021  
Db 7006 SVHLT-----WEPDDGGSPLTGYVVEKREVSRTWTWKVMDFTDLEFTVFDLVQGGKEYL 7061  
Qy 5022 -----SPAEVDELEFLSADSGRAEPPEPADWQYRDEHFI----- 5058  
Db 7062 FKVCARNKCGGEFAYV--DEPNWGSTPATVDPDENVKWORTANSIFLTDWPPKNDGGS 7120  
Qy 5059 -----CIR-----FEALTEARQAVTRFOE--MFATLIGIGVEIKLVEGPPRVE 5099  
Db 7121 RIKGYIVERCPRGSKWACGEPVAETKQVETGLEEGKWA-----YRVK 7165  
Qy 5100 MCISKETAPVVPPEPLSLTSDAAPVFLTE--LQOQEVODGYVPFSDCVVTCQPMPSV 5157  
Db 7166 ALNRQGASKRPRTEBIOQAVDTQEAPEIFLDVKLAGUTVRAGTKIELPATVTCQPEPKI 7225  
Qy 5158 RWFKDGKLEEDDHYMINEDQGGHQLIITAVPADMGVYRCLAEANGVSGSSTKAEALRVD 5217  
Db 7226 TWTKADMILKQDKRITI--ENVPKSSTVIVDSKSDTGTYYIEAVNVCGRATAVVEVNV-- 7283  
Qy 5218 LSTSDYDTAADATSSSVFSAGYLSREQGTSTTDEGQLQPVVEELRLQVAPGTRL 5277  
Db 7284 LDKFGPPAFTDVTNESCILLTNWPPRDDGGS----- 7316  
Qy 5278 AKFOLKVKGYPAPRLYKFKDQPLTASAHIMTKGKILHTLEIISVTRSDGQVAAVSN 5337  
Db 7317 -----KINYVVER-----RATDSEVWHKL--SSTVKDINFKATKLIP 7352  
Qy 5338 AMGAAYSSARLLVRGDFPEKPSADSVHQLVPRMLERFTPKVKXKGS--SITESVKVE-- 5395  
Db 7353 NKEYIFRVAANMVGVPVQASPIAKYQDPDPGPPTRLEPSSDITKDAVTLTWCPEPDD 7412  
Qy 5396 -GRPVTVHMLREAEERGVLMIGDPTPGTYTVASSAQHSLVLLDVGRHQHTY--TCIAS 5452  
Db 7413 GGSPI--TGIVW--ER-----LDPTDKWRCNMKVPKDTYRVKGLTNKKYFRVLA 7463  
Qy 5453 NAAGQALCASLHVSGLP--KVEEQEKYKXALISTFLOQ--TQOISAGQLETSFADLGG 5509  
Db 7464 NLAG-----PGKPSKSTEPILIKDIPDPWPPGKPTVKDVGKTSVA-----LNW 7507  
Qy 5510 ORKEEPLAAKALGHLSLAEVGTTEFLQKLTSQITEMVSAKITQAKLQVPG--GPSDEDSK 5568  
Db 7508 TKPEHDGAKIESYVIEMLKTGTDEWVR-----VAEGVPTTQHLPLGMEGQBYSF 7558  
Qy 5569 TPSAPRGRGRSPSSIOESSSESDGDARGEIPIYVVTADYPLPLGAEQDAITLREGQY 5628  
Db 7559 RVRAVNKAGESEPSE-----PSPDVLCKEKL 7585  
Qy 5629 -----VEVLDAH--PLRWLVTKP-----TKSSPRQGW--VSPAYLDRRL 5666  
Db 7586 PPSPPRLVLEVINITKNTADLKWTPPEKDGSPITNIIYVEKEDVRKQWQVDTVTVKOTKC 7645  
Qy 5667 KLSIP--ENG-----AAEAPFPGGAUSEDEYKARLSSVIOELLSSEQAFVEELQFLQSH 5718  
Db 7646 TVTPTLSEGLYFRVAAB-----NAIQSDY-----TEIEDSVLAKDTFT----- 7695  
Qy 5719 HLOHLERCPHPIAVAGOKAVIFRNVDRIGRHFSLQELQOCDTDDVAMCFI--KNQAA 5777  
Db 7696 -----TPGPPYALA-----VVDVTKRHVDLKWPEPKNDGGRPIQRYVIEKKERL 7729  
Qy 5778 FEQYLEFLVGRV-----QAESVTVSTAIQEFYKKAIE-----EALLAGDPS 5818  
Db 7730 GTRWTK--AGKTAGPDCNFRVTDVIEGTEVQ--FOVRAENAGVGHSPSEPTIELSIEDPT 7785

Qy 5819 QPPPPLOHYLE-----QVVER-----VORYQALLKELIRNK 5850  
Db 7786 SPSPPLDLUHVTDAGRKHIAIAWKPEKGGSPITGYHVEMCPVGTETKMRVNSRPIKDL 7845  
Qy 5851 ARNRQCALLLEQAYAVVSALPQRAENKLHVS--LMENYPG-----TLSEALGEPPIQ 5899  
Db 7846 KPAVBEGVVPDKREY-----VLRVRAVNAIGVSPSEISENVVAKDPCKPTID--LET 7896  
Qy 5900 GHPIWEG-----APGARMPWKGNRHV-----FLFRN-----HLVICKP-RED 5937  
Db 7897 HDLIIVTEGKLSIPVFRAPVPTVSMHKGKVEKASDLRTWKNDHISAHLEVPKSVRAD 7956  
Qy 5938 SRDITVSYVRNMKLSISDLNDQVEG-----DDRAFEV-----WQ----- 5973  
Db 7957 AGIYITIT--LENKLGSAIASINVKVLGPGCKDIKASDITKSSCKLTWPEPEFDGGTPI 8014  
Qy 5974 -----BRESVRKYLLOARTAIKSSW--VKEIC-----GIQORLALPV 6010  
Db 8015 LHVLERREAGRTYIPVMSGENKLSWTVDKILPNGEYFRVRAVKNVKGGEYIELKPV 8074  
Qy 6011 -----WRPDP--FEEELADCTAELGETVKLACRVGTGPKPVISWYKDGKAVQVDPHHIL 6063  
Db 8075 IAQDPKQPPDPVDVEVHNPTAE-----AMITWKPPL--YDGGSKIM--CYIIE 8120  
Qy 6064 EDPDGS-----CALILSLTGVDSG--QYMCFAASAAGNSTLIGKILVQVPP-R 6109  
Db 8121 KIAKGBERWKRGNELHVPILTYTAKGLEEGKEYQFVRAENAAG--ISEPSRATPTPK 8176  
Qy 6110 FYNKVPASFPV-----EGEDAQFTCTIEGAPYQIRWYKD-----GALLT 6149  
Db 8177 AVDPIDAPKILLTSLEVVRGDEIADASISGSPYPTITWIDENVIVPEEKRAAPLV 8236  
Qy 6150 TGNKFO-----TLSEPRGLLVIRAAASKEDLGYECELVNRGLSARAS 6194  
Db 8237 RREKGEVQBEPPVPLTQRLSIDNSKGESQLRVSLSRDPHGLWIKVENDHGIKAP 8296  
Qy 6195 AELRIOS-----PMLQAOEQCHREOLVAVE-----DTTLERADQE-----V 6231  
Db 8297 CTVSVLDTGPPINFVFDIRKTSVLCKWPEPLDDGSGSIINYLEKKDKTKPDSEWIV 8356  
Qy 6232 TSVLK-----RLLGPK-----APGPS-----TGDLTGPGPCP----- 6258  
Db 8357 TSTLRHCKSVTKLISGKEYLFRVRAENRFGPCPCVSKPLVAKOPFGPDAPDKPVED 8416  
Qy 6259 -----RGAPAL-----QETG----- 6268  
Db 8417 VTSNSMLVKNWEPKNGSPILGTYLKEKREVNSTHWSRVNKSLLNALKANVDGLLEGLTV 8476  
Qy 6269 -----SOPPVGTSEAPAVPPVPPQPLHEGPECEPEAIARAQEWTPIRM 6314  
Db 8477 FRVCAENAAAGPKFSPSPDKTAHDPISPGPPIFRVTSTSTTEL-----EWEPPAPN 8531  
Qy 6315 EGAAMPG-----ACTGELLWD-----VHSHVVRETTQ--RTVYQAIPTHTA----- 6354  
Db 8532 GGGEIVGYFVDKQLVGTNE--WSRCPEKMKVRYQVTKREIBEGADYKLRVASVNAAGEGP 8589  
Qy 6355 -----RPSMQVTIE--DYQAOQTGTAQFALEGGDPQSPVWYKDSQVLV 6398  
Db 8590 PGQTQVTVAEPOEPFAVELDSVKGGIQIMAGKTLURIPAVVTGRVPVTKVTKESGEL- 8648  
Qy 6399 DSTRLSQQOEGTITYSLVRHVASKAGVYTCLAQNTGGQVLCKAELLVLGGNPE-----D 6454  
Db 8649 DKDRVIDNVGTKSELI--KDALKRKHGRVITATNCSGSKGFAAARVEVE--DVPGPVL 8705  
Qy 6455 SEKQSHRRK--LHSFVYEKBEIRGVGFVKRVQHKGNKILCAAKFIPLSRTRAQAYRE 6512  
Db 8706 LKPVVTRNRKMLNWSDDPDGSGSEITGFI--IERKDAKN----- 8743  
Qy 6513 RDILAALSHPLVTGLDQFETRKTLLILELCSSELDRLYRKGVVTEAEVVKVYQOLV 6572  
Db 8744 -----HTWQPIETERS--KCDITGLL 8763



QY 6573 EGLHYLHSHGVL-----HLDIKSNILMVHPAREDIKIDFGFAQNIITPAELOFSQ 6623  
 DB 8764 EGQEY--KFRVIANKKFCGPPVIGP--ILAVDP-----8794  
 QY 6624 YGSPFVSPFELIQNPVSEASDIWAMGVISVLSLTCSPFAGESDRATLLNVLEGRVSW 6683  
 DB 8795 LGPP--TSPE-----RLTYETKST-----ITLDWKEPRNGG 8826  
 QY 6684 SPMAHLSE---DAKDFIKATLQAPQAPSAQCLSHPWFLKSPAEBAHFINTKOLK 6739  
 DB 8827 SPIQGYIIEKRHRKDPD-----ERNKRLCPTTSLVENLDEHQMYEPRVK--- 8873  
 QY 6740 FLLARSRQWSLMSYKSLVMWESIPELLRGPDPDSGLGVARHLCRDTCGSSSSSSSDNE 6799  
 DB 8874 -----AVNEIGSEPSLP-LNVV-----IQDDE 8895  
 QY 6800 LAPFARAK-----SLP---PSFVTHSLP-----LHPRGFLRPSASL---PEE 6835  
 DB 8896 VPPTIKLSLVRGDTIKVKAGEPVHVPADVTGLPMPKIENSKNETVIEKPTDALQITKEE 8955  
 QY 6836 AEASERSTEAPASPAGCPGAPPAQCGVPRHSVIRSLFYHQAGESPEHGALAPGSRHP 6895  
 DB 8956 VRSRAKTESLIPKAVREDKDTYTVTASNRGLGSVFRNVHVEYDR-----PSPPRNL 9007  
 QY 6896 ARRRHLLKGGYIAGALP---GLREPLMEHVRVLEEAAREEOA-----TLLAK----- 6939  
 DB 9008 AVTDIKAESCVLTWDAPLDNGSE--ITHYVIDKRDASRKAKEVEEYNTAVEKRYGIWK 9065  
 QY 6940 -APSFETALRLPASGTHLAPGHSLSLHDSPTSPRPSSEACGEAQRLPSPAGPIRDM 6998  
 DB 9066 LIPNGQYFRVRAVVKY---GISDECKSDKVIQDP-----YRLFGPP--GKP-KVL 9111  
 QY 6999 GHQOSKOLPST---GGHPGTAQ--PERPSPDSP-WGQAPFPCHPKQG-----SAPO- 7044  
 DB 9112 ARTKSMVSWTTPPLDNGGSPITGYWLEKREGSPYWSRVRAPITKVGLKGVEFNVRL 9171  
 QY 7045 -EGCS-PIHPAVAPCPGFPFGSCKEAPLVPSPP-FIGQFOAPPAPAKASPPILDSKXGPG 7101  
 DB 9172 LEGVKYQPRAMAINAAGIPSE-----PSDEVAGDPFIFFPGP--PSCPEVKDKT-KS 9222  
 QY 7102 DISLPGRPKPCSPGASQASQSSQVSSSLVGVSSGVGTGPGSLDAEGW--TOEADLS 7159  
 DB 9223 SISLWKP---PAKDG---SPKGYIVNQEEGT-----TDWGRVNEPKLI 9264  
 QY 7160 DS---TPTLQRPQEQATMRKXFLGGRGYAGVAGYGTFAFGDAGMGLQGPGPWARIW 7215  
 DB 9265 TTCCEVWNLKE-----LRKYF-----RVK- 9285  
 QY 7216 AVSQSEEEQEARAESQEQEARABSPLOVSARPVFVG-----RAPT 7262  
 DB 9286 AVNEAGESPDDTGTETPATDQEE-----PEVFI---DIGAQDCLVCKASQIRIPA 9335  
 QY 7263 RSSPEPTP---WEDIGQVSLVQ---IRDLSDAADAADTISLIDSEVPAYNLNLDYDI- 7315  
 DB 9336 VIKGPTKSWEPDGRKAKWKGVDHDPDAQLEAENSIVIIIECKRSHTKYSIT 9395  
 QY 7316 -KYLFFEMIRKVPKSAQPPSPMAEELAEFP-EPTWMPGELG-----7360  
 DB 9396 AKNAGQKANTACRVKVMVDPGPPKDLKVSITRGSCLRSKWMKMPDDGGRKIGYVIEKRT 9455  
 QY 7361 -----PHAGI-----EITESEEDVDALLAAAGVKRKSPPSLSLFFPGRHL 7404  
 DB 9456 IDGKANTKVPDCSTTFVWDLSEQQYFRVRAENRFGIGPPVETIQNTARDP--IY 9513  
 QY 7405 PLDSPAELGLRBRVKAQV-----EHSIRLKGPEGLEKEG-----7440  
 DB 9514 PPDPPKILKILITKNTVHLSWKPKNDGGSPVTHYIVECLAMPDPTKKEAWRQCNRD 9573  
 QY 7441 -----PPRKPCLASFRLSGLKSM-----DRAPTF-LAELESDTY 7474  
 DB 9574 VEELQFTVEDLVEGGEYEFVRVANAAGVKSPSATVGFQDQCRDMPSPSIDLKEFWE--V 9631  
 QY 7475 VLGSVTLACQVSAQPAQAATWSKDGAPLESSRVLISATLKNFOLL--TILVVVAE--- 7529

DB 9632 ESGTNVNIIVAKIKGVFFPTLTWFKAPPKPKPKDKEPVLYDTHVKNLVDDCTCTLVIPQRR 9691  
 QY 7530 -DLGVYTCVSNALGTVTITGVLRKAERPSSPCPDIG-----EYVADGVLLWKPVESY 7583  
 DB 9692 SDTGLYITAVNNLGTASKEMRLNLGRFG---PPVGPVKFESVSADQMTLSNFFPKDD 9747  
 QY 7584 GP---VTIVQV-CSLEGGSWITLASDIFDCCVLSKLSRGGTYTFTACVSKAGNG-PYS 7638  
 DB 9748 GSGKITNYVIERKREANRKTWVHVSSEPKECTYITPKLLEHGYVFRIMAQNKYGIGEPLD 9807  
 QY 7639 SPSE 7642  
 DB 9808 SEPE 9811  
 RESULT 8  
 Q10466 PRELIMINARY; PRT: 26926 AA.  
 AC Q10466;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DE Titin, heart isoform N2-B (EC 2.7.1.-) (Connectin).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=HEART;  
 RX MEDLINE=96026330; PubMed=7569978;  
 RA Labelit S., Kolmer B.;  
 RT "Titins: giant proteins in charge of muscle ultrastructure and  
 RT elasticity.";   
 RL Science 270:293-296(1995).  
 [2]  
 RN SEQUENCE OF 22277-25376 FROM N.A.  
 RP MEDLINE=9225380; PubMed=1582406;  
 RA Labelit S., Gautel M., Lakey A., Trinick J.;  
 RT "Towards a molecular understanding of titin.";   
 RL EMBO J. 11:1711-1716(1992).  
 [3]  
 RN SEQUENCE OF 1976-2014 FROM N.A.  
 RP Labelit S.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RN CHARACTERIZATION.  
 RP MEDLINE=9533114; PubMed=7607248;  
 RA Gautel M., Castiglione-Morelli M.A., Pfuhl M., Motta A., Pastore A.;  
 RT "A calmodulin-binding sequence in the C-terminus of human cardiac  
 RT titin kinase.";   
 RL Eur. J. Biochem. 230:752-759(1995).  
 CC -!- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE  
 CC ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF  
 CC SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.  
 CC -!- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.  
 CC -!- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE  
 CC KINASES.  
 CC -!- IMMUNOGLOBULIN C2-LIKE DOMAINS AND 132 FIBRONECTIN TYPE III-LIKE  
 CC DOMAINS.  
 DR EMBL; X64698; CAA45939.1; -  
 DR EMBL; X83270; CAA58243.1; -  
 DR EMBL; X64697; CAA45938.1; -  
 DR EMBL; X90568; CAA62188.1; -  
 DR EMBL; X64699; CAA45940.1; -  
 DR FIR; I38344; I38344.  
 DR PDB; 1BPV; 12-AUG-99.  
 DR PDB; 1GIC; 12-OCT-01.  
 DR PDB; 1NCT; 08-NOV-96.  
 DR PDB; 1NCU; 08-NOV-96.  
 DR PDB; 1TIT; 11-JUL-96.



DR PDB; 1TTU; 11-JUL-96.  
 DR GO; GO:0005856; C:cytoskeleton; IEA.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; P:ATP binding; IEA.  
 DR GO; GO:0005516; F:calmodulin binding; IEA.  
 DR GO; GO:0004896; F:hematopoietin/interferon-class (D200-domain. . .); IEA.  
 DR GO; GO:0004601; F:peroxidase activity; IEA.  
 DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.  
 DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO; GO:0007511; P:muscle development; IEA.  
 DR GO; GO:0004668; P:protein amino acid phosphorylation; IEA.  
 DR GO; GO:0006979; P:response to oxidative stress; IEA.  
 DR InterPro; IPR000282; Cytok\_receptor\_2.  
 DR InterPro; IPR000577; FGK Kin.  
 DR InterPro; IPR003961; FN III.  
 DR InterPro; IPR008957; FN III-like.  
 DR InterPro; IPR001092; HLH Basic.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR002016; Peroxidase.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR008266; Tyr\_pkinase\_AS.  
 DR Pfam; PF00041; fn3; 132.  
 DR Pfam; PF00047; Ig; 91.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00060; FN3; 127.  
 DR SMART; SM00408; IGc2; 23.  
 DR SMART; SM00220; S\_TKC; 1.  
 DR PROSITE; PS00933; FGK\_KINASES\_1; 1.  
 DR PROSITE; PS00038; HLH\_1; 1.  
 DR PROSITE; PS08335; IG LIKE; 87.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 DR PROSITE; PS00435; PEROXIDASE\_1; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW Muscle protein; Cytoskeleton; Structural protein; Calmodulin-binding;  
 KW Serine/threonine-protein kinase; Alternative splicing; Repeat;  
 KW Immunoglobulin domain; Phosphorylation.  
 FT DOMAIN 1370 1389  
 FT DOMAIN 4429 4614  
 FT DOMAIN 24731 25070  
 FT DOMAIN 25030 25056  
 FT MOD\_RES 1372 1372  
 FT MOD\_RES 1377 1377  
 FT MOD\_RES 1382 1382  
 FT MOD\_RES 1387 1387  
 FT MOD\_RES 26171 26171  
 FT MOD\_RES 26178 26178  
 FT MOD\_RES 26184 26184  
 FT MOD\_RES 26190 26190  
 FT MOD\_RES 22277 22277  
 FT CONFLICT 22449 22449  
 FT CONFLICT 22454 22454  
 FT CONFLICT 23324 23324  
 FT CONFLICT 23324 23324  
 FT SEQUENCE 26326 AA; D5E6CD3254DF5523 CRC64;  
 Query Match 8.7%; Score 3577.5; DB 4; Length 26926;  
 Best Local Similarity 20.2%; Pred. No. 4.6e-181;  
 Matches 1929; Conservative 1288; Mismatches 3662; Indels 2663; Gaps 320;  
 QY 4 POFSGAPFLTPKAFVSVGKDALTSQIVGNPTPQVSWKQDQPVTAGAFRLA---Q 60  
 DB 1030 PGEAPAPFTRPVQKLVGEGSVVFGVGVGNPKPHYVWKSGVPLTTGYRYKSVNKK 1089  
 QY 61 DGDLYRLTLDLALGDGGYVCRARNAIGFAAVGL-----QVD 100  
 DB 1090 TGECC-KLVISMTFADAGEYTIIVNKHGEISASASILLEADYELLMKSSQEMLYQVTV 1148

QY 101 A---EAACAEQAPHF----- 113  
 DB 1149 AFVQEPVGTATPGFVYSEYEKEQALIRKKMAKDTVVVTVYVEDQEFHISSEERL 1208  
 QY 114 -----LRPTSIRVREGSEATPCRV 133  
 DB 1209 IKEIYRIIKTTLEELLEDGEEKMAVDISESAVSGFDLRKIKRYRIILEGMGVTHCM 1268  
 QY 134 GGSPPRAVSWSKDGRRLGEPDGPVRVVEELGEASA-LRRAAPRPGGTYVEVAENPLGA 192  
 DB 1269 SGYPLPKIAWYKDGKLI--KHGERYQWDFLODGRASLRFPVLPEDEGYTFASNIKGN 1326  
 QY 193 ASAAALVVDSDAADTASR--PCTSTAALLAHLQRR--REAWRAEGAFASPP----- 240  
 DB 1327 AICSGKLYVEAPAPLGAFTYPTLEPVSIRISLSPRSVSRSPIRMSPARMSPARMSPARM 1386  
 QY 241 -----STGTR-----TCTVTEGKHARLSCVVTGEPKPTVWK 273  
 DB 1387 SPARMSFGRLEETDESQLERLYKPVVLKPVSKLEGANCFDILKVVGRMPETFWFH 1446  
 QY 274 DGQLVTEGRRH--VYVEDAQENFVLKILFKOSDRGLYTCTASNLVGYQTVSSVLVV-- 328  
 DB 1447 DGQIVNDYTHKVIKEDGTQSLI--IVPATPSDSEWTWVAQNRAGRSSIVILVEAV 1504  
 QY 329 ----- 328  
 DB 1505 EHQVPMFVEKLKNNVNIKEGSRLEMKVRATGNPNPDIVLKNSDIIVPHKYPKIRIEGT 1564  
 QY 329 -----R 329  
 DB 1565 GEAAALKIDSTVSQDSAWYTATINKAGRDTTRCKVNVVEFAPEPERKLIIPRGYRAK 1624  
 QY 330 EPAVP-----FKRLQDLEVRKESEATFCE---VP 357  
 DB 1625 EIAAPELEPLHLRYGOEWEGDLYDKKQKQFFKKKLTSLRLKRGFAHFCRLTPIS 1684  
 QY 358 QPSTEAAPKEETRL-----WASAKYGI-----EEEGTERRLT 390  
 DB 1685 DPTWVWELHDGKPLEAANRLMINFPGVCLSDYGVAYSRDSGIITCRATNKYGTHTSA 1744  
 QY 391 VRNVSDADDVAVICETPEGSR-----TVAELAVQGNL-----LRKLPRKT 430  
 DB 1745 TLIVKDEKSLVBSQLPEGRKGLQRIEELERMAHEGALGVTTDQXKQKPDVILYPEPV 1804  
 QY 431 AVRVGDTAMP-CVELAVPVPVHMLNQEVVAGVAISAEGRHTLTISQCLDVGQ 489  
 DB 1805 RVLEGETARPCRVTVGPQPKVNNYLNGLQIRKSKFRVRYDGI-HYLDIVDCKSYDTGE 1863  
 QY 490 VAFWAG-----DCQTSTRECVSAPRKPPLQ-----PVD 518  
 DB 1864 VKVTAEINPEGVIEHKVLEIQQRDFRSVLRRAPEPRPEFHVHEPGKLOFEVQKVRPVD 1923  
 QY 519 -----PVVKARMSSVILSWSPPHGE-----RPVTIDGY-----LVEKKLGTW 560  
 DB 1924 TTETKEVVKLRAERITHEKVPESSELRKSKFRTEEGYEAITAVELKSKKDSYEE 1983  
 QY 561 I----RCHAEAWVATPELTVD---VAEEGNFOFRVSALNSFCQSPYLEPFGPVHLPK 613  
 DB 1984 LLRKTKDELLHW--TKELTEEEKKALAEKGIPTTFKPKIELSFSE-----APKI 2034  
 QY 614 AVTPLKAVQAVGEGVTVFSDLTVASAG-----EWFLDGQALKASS-VYEIHCORTRH 666  
 DB 2035 FERIQSQTVG--QGSDAHFRVRV-----VGKPOPECEWYKNGVKIERSDRIYWPEDNVC 2088  
 QY 667 TLITREVPASLHQAQLKF-----VANGIESIRMEVRAAPGLTANKPAPAAAREVLARLHE 722  
 DB 2089 ELVIRDVTAIE-DSASIMVKAINTAGTSSHAFLVQAKQILITFTQ-----ELQDVVAKEND 2143  
 QY 723 E-AQLIAELSDQAAAATWLKDGFTLSFGPKYVQASAGRRLVLRDVARDDAGLIECV-- 779  
 DB 2144 TMAFTCETSEPFVKVKKWYKDGMEVHEGDKYRMSDRKHVHFLSILITDTSDAEDYSCVLV 2203



Db 4271 EVNWTFENKLVPSDEKFKCLODQNTYTLVLDKXNTEDHQGEYVCEALNDS-----GKTAT 4325  
Qy 2459 SARL-VWSDIPVILTRPLEKPTGRELQSVLWCDPRPAPKA-VOMYKDDTPLSPSEKPKM 2516  
Db 4326 SAKLTUVKRAAPVIRKRIEPLVALGHLAKFTCBQIAPNVRVQWFRAGRIEYSDKCSI 4385  
Qy 2517 SLEGOMAEIRLRLMPADAGVYRQAGSAHSSTEVTVAREVTVTG-----PLQDA 2567  
Db 4386 RSSKYISSLERLQVQVDCGYTCKASNEYGSVCTA-TLTVTVPGGKKVRKLLPERKP 4444  
Qy 2568 BATEGMAFSCELSHDEEVEWMSLNGMPLNDSFHEISHKGRHRTLVLKSIQRADAGIV 2627  
Db 4445 EPKEE--VVLKSVLRKRPEEPEKPEPKL-----EKVKKPAVPEPPPPKPEVEVEPTV 4497  
Qy 2628 RASSLKVSTARSARLEVRVVPVFLKALDLDLSABERGTLALQCEVSDPE-----AHV 2678  
Db 4498 TKRERKIPEPTKVP-EIKPA:PLPAPEPKPEAEVKTIKPPPEPBPPTIAAPVTVFV 4556  
Qy 2679 WRKQVQLGPDVKYDFLHTAG-TRGLVVHDVSPDAGLYTCHVGS-----EETRARVRV 2731  
Db 4557 GKABAK--APKEBAAPKPKPIKGVPKTSPIEAERKLRPGSGGKPPDEAPFTYQL 4613  
Qy 2732 HDLHVIGITKRLKTMVELE----GSCSPCEVLSHESADPANWTVGKTVGSSRFOATR 2787  
Db 4614 KAVPLKFKYKEIKDIILTESEFVGSAPFECVLSPSTAI--TWMKDGNSIRESKHFIA 4671  
Qy 2788 QGRKYLVRRAAPSADAGEVVFVSR---GLTSKASLIVRERPAAIKPLEDO-WVAPGE 2842  
Db 4672 DKORKLHIIIDVLSADAGEYCVLRLGNKEKSTAKLVEELPVRFVKLTLEEVTVVKGQ 4731  
Qy 2843 DVELRCELSRAGTPVHWLKDRC-AIRKSQYDVVCEGTMAMLVIRGASLKADAGTYCEVE 2901  
Db 4732 PLYLSCELNKE-RDVVRKDGKIVVEKPGRIVPGVIGLMLRALTINDADTDAGTYTVTVE 4790  
Qy 2902 ASK-----STASHVEEKANCETEELTNLOVEKGTAVFTCKTEHPAATVTRKGLLELRA 2957  
Db 4791 NANNLECSCKVKEVIRDLVKPIRDQHVXPKGTAFACDIADKTPNIKFKGYDEIPA 4850  
Qy 2958 --SGKHQPSQEGTLRLTALITSALEKADSDTYTCDIGQAQSRACLIVQGRVHIIIDLEDVD 3015  
Db 4851 BPNDKTEILRDGNHLYLXIKNAMPEDIAEYAVEIGKEYPAKLTJGEEVEVLLKPIEDVT 4910  
Qy 3016 VOEGSSATFRCKISPAVPEVPHWFLDKTPLHANELNEDAOGGVHVLTLBQALAKSGT 3075  
Db 4911 IYKESASFDAISEADI-PGOWKLKGELLRSPPTCEIKAE-GGKRFLLTHKVLDDQAGE 4968  
Qy 3076 IYFEAGDQASAAARVTEKPSVFSRELTDATITEGEDLTLVCETSTCDIPMCWTQDKTIL 3135  
Db 4969 VLYQALNALTALLTVKIELDFAVPLKDVTPERRQARFEC-VLTREANVWNSKGPDI 5027  
Qy 3136 RGSARCOLSHEGHRACQLLITGATLQDSGRYKCEAGGACSSIRVVRHAPVFPBALKDLE 3195  
Db 5028 XSSDKFDIADGKHILVINDSQDFDEGVYTAEEVEGKTSARLFTVGTIRLAFMSPLEDQT 5087  
Qy 3196 VLEGGAATLRCVLSVAAPVKCYGNVLRPGDKYSLRQEGAMLELVNLRNLRPODSGRYS 3255  
Db 5088 VKEGETATFVCELSHEKMHVWFKNDAKLHSTRVTLISSEGTHKLEMKFETLDDISOIK 5147  
Qy 3256 CSFGDQTTSATLVTALPAQFICKLRNKEATGATATLRCLESTKAPVSWRKGSETLRDG 3315  
Db 5148 AQVKELSSTAQLKVLLEADPYTVKLHDKTAVEKDITLKCEYSKDVVPVKFKDGEIIVPS 5207  
Qy 3316 DRYCLRODAMCELQIRGLAMVDAEYSCVCEERTSASLTIRPMPAHFIRGLRHEQESIE 3375  
Db 5208 PRYSIKADGRILKIKKADLKDGEYVCDGCTDKTKANVTVEARLEIIVEKPLYGVEVVF 5267  
Qy 3376 GATATLRCELS-----KAAPVWRKGRSRLRDGDRHSRLQDAGVCELQICGLAVA 3425  
Db 5268 GETAHEIELSFPDVHGQWKLKQPLTASPDCEIIEDGKX-----ILLHNLQCLM- 5319  
Qy 3426 DAGEYSCVCGEBRTSATLTKVALPAKFTTEGLRNEEAVEGATAMLCELSKVAPV-EWRKG 3484  
Db 5320 -TGEVSFOANAKSANLKVLEPLTIFITPLSDVKVFEKDEAKFECEVSREPFTFRWLKG 5378

Qy 3485 PENLRDGRYIILROBGTRCCELOICGLAMADAGEYLVCVCGQERTSATLIRALPARFIEDV 3544  
Db 5379 TOBITGDDRFELIKDGTGHSVMVKSAAFEDEAKYMFEAEDKHTISGKLIIEGRLRFLPTD 5438  
Qy 3545 XNCEAREGATVLCQCL-NSAAPVWRKGSERLRDGRYSRLQDGTCKCELOIRGLAMADT 3603  
Db 5439 KDVTAKEKESAVFTVELSHDNLRVKNQDRLHTTRSVMQDEGKHTSITFKDLSIDT 5498  
Qy 3604 GEYSCVCGQERTSAMLTVRALPIKFTTEGLRNEEATGATVRLCRLSKM-APVWKKGHE 3662  
Db 5499 SQIRVEAMGMSAEKTLVLEGDPYFTGKLQDYGVEKDEVILOCEISKADAVKVPKQOK 5558  
Qy 3663 TLRDGRHSRLRODAGARCELQIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPKFTLEGLN 3722  
Db 5559 EIKPSKNVITDKGKRMJLLKALKSDIGQYTCOGTDTKSGKLDIEDREIKLVRPLHS 5618  
Qy 3723 EE----- 3724  
Db 5619 VEWNETARFETEISEDDIHANWKLKGEALLQTPDCEIKBEGKIHSVLVHNCRLDQTCG 5678  
Qy 3725 -----ATEGDTATLWCELS-KAAPVWRKKGHEIL 3752  
Db 5679 VDFQAAVWKSASHLRVKPRVIGLLRPLKDVTTAGETATFDCELSYEDIPVWYLGKKL 5738  
Qy 3753 RGDHRSRLQDGRSCELQIRGLAVWDAGEYSCVCGQERTSATLTVRALPARFIEDVKNQE 3812  
Db 5739 EPSDKVVRSEKGVHTLRLDVKLEDADEGVQLTAKDFKTHANLFVKEPPEVFTKPLEDQT 5798  
Qy 3813 ARGATAVLQCELSKA-APVWRKGSERLRDGRYSRLQDGTTCRCELOIRGLSVADTGEYS 3871  
Db 5799 VEGATAVLECEVSRNENAKVKNFKNTEILKSKYIEIVADGRVRLKVLINDCTPED:KTYT 5858  
Qy 3872 CVCGQERTSATLTVRAPQVPFREPLOSQAEGESTATLOCELSEPTATVWMSKGGJLQQA 3931  
Db 5859 CDAKDKFTSCNLNVVPHVEFLRPLDLQVREKEMARFECSELSRENKVKWFKDGAIEIK 5918  
Qy 3932 NGRREPLQCTAEVLVLDQREDTGEYTCOSQATSATLTVTAAPVRLRLOQOEVD 3991  
Db 5919 GKRYDIISKGAVILWINKCLLDDEAEYSCEVTAETSGMLTVLEBEAVFTKLANIEVS 5978  
Qy 3992 EGCTAHLCCSELSRAGASVWRKGSLOLFPCKAYQMVQDGAALVLRGVEQEDAGDYTCD 4051  
Db 5979 ETTIKLVCEVSPGAEVWYKGDDEIITGTYEILTREGKRLTVIONAHLSDAGNYNCR 6038  
Qy 4052 TGTQSMASLSVRVPRPKFTRLQSLQEQTGDIARLCCOLSDAESGAUVOWLKEGVELHA 4111  
Db 6039 LPSRTDGVKVKHAAAEFISKPNLEILEGEKAEFVCSIS--KESFPVQWKREDDKTLES 6096  
Qy 4112 GPKYVRSQATRELLIHOLEAKDTGECYACVTGGQXTAASRLRVTPEVTIIVRGLVDAEVT 4171  
Db 6097 GDKYDVIADGKRVLVVKOATLQDMGTYYVWGAARAAHLTVIE-KLRIVPLKDRVK 6155  
Qy 4172 ADEDFEFSCEVSRAGATGVQWCLQGLFLOSNEVTEVAVRDGRIHTLRLKGVTPEDAGTVS 4231  
Db 6156 EQGEVVFNCENVTEGAKAKWFRNEEAI:FDSSKY--IILQKDLVYTLIRDAHLDDQANYN 6213  
Qy 4232 FHLGNH-----ASSAQITVRAPETVILEPLQDVLQSEGODASQCRLSRASGOEAWLG 4286  
Db 6214 VSLTNHGRNVKSAANLIVVEEDLRIPELKDITETWEKKSVTFCWKVNRNL-VTLKWTKN 6272  
Qy 4287 GVPLQANEMDITVEOQTLLHLLTHKVTLEDACTVSHVGTCSSEAKLK-VTAKNTVVRG 4345  
Db 6273 GEEVFPDNRVSYRVDK-YKHMJTIKDCGPDDEGEYIVTAGQDKSVAELIIEAPTEFVEH 6331  
Qy 4346 IENVEAL:EGEALFECOLSOPEVAHTWLLDDPEVTSNAEVVFPENGRLHLLILKNLR 4405  
Db 6332 LEDQTVTEFDVAFSCQLSR-EKANYKVRNNGREIK--EGKTKFKFGDSIHLIITKDCR 6388  
Qy 4406 PQDSRVTFFLAGDMVTSARLTVRGWRLILEPLKNAAVRAGAQAARCTCTLSEAVPVGEAS 4465  
Db 6389 LDDECEYACGVEDKSRARLFEIPEIIRPPQDILEAPGADVFLAELNK--DKVEVQ 6446

QY 4466 WYINGAAVQPDSDWTVTADGSHOALLRSQAOPHHAGBVTACRDVASAR----- 4516  
Db 6447 WLNNMNVVQGDKH-QMMSEKIHRLQCDIKPRDQGEYRFIADKGEARAKLEAAAPKI 6505  
QY 4517 ----- 4516  
Db 6506 KTADQDLVVDVKPLTWVPYDAVPAKAEWFKENEPLSTKTDITAEQTSFRILEAKKG 6565  
QY 4517 -----LTVLGLPDPPEDEAVVAHSHHTVLSWAAPMSDGGGLC 4555  
Db 6566 DKGRYKIVLQNHKGKAEQFINLKVIDVPGVVRNLEVTTFDGEVSLAWEELTDGGSKII 6625  
QY 4556 GYRVEKGCATQWRILCHELVGECVVDGLAPGET-YRFRVAAGVPVGAPEVHLPTQV 4614  
Db 6626 GYVVERRDIKRTWVLAIDRAESCEFTVGLQKGVYELFRVARNRVGTGEPVETDNPV 6685  
QY 4615 RLAEPPKVPPOP-----SAPESRQVAGEDVSLEVVABAGEVIMHKGM 4660  
Db 6686 E-ARSKYDVPGLNLVTITDVRFGVSLTWEPPEYDGAETINYVIELRDKTISRWDTAM 6744  
QY 4661 ERIQPGRFVYVSGROQMLVIKGTAEQGEYHCGLAQGSICPAATFOVALSPASVDE 4720  
Db 6745 T-----VRAEDLSATVDVSGQYSFRVRAQNRIGVK-----PSAATPVKVADPIERP 6795  
QY 4721 APQPSLPPEAQEGDLHLWLALARKMSREPTLDSISELPBEDGRSQRLOPQAEVAP 4780  
Db 6796 SPPVNLTSDDQTQSSVQKWBPLKD---GGSPILGVIIERCEE-GKDNWIRCNM-KLVP 6850  
QY 4781 DLSEGYSTADELARTGDADLSHTSDDESAGTSLSVTLKKGACRPGTSLASKVGAPAA 4840  
Db 6851 ELTYKVTGLEK-GNKYLVRVAENKAGVSDSEI----- 6883  
QY 4841 PSVKPQOOQELAAVRPPLGLDSTKD-LGDPMSMDKAAVK-----IOAAFKGYK 4887  
Db 6884 -----LGLTADDAFVPTMDLSAFKDGLEIVPNPITILVPSTGY- 6924  
QY 4888 VRKEMKQEGPMFSHTFGDTAQVDALRLCVCVASADVRARWLDGVELTDGRHHID 4947  
Db 6925 -----PRPTATWCFDGVLETDGDRVOMKTLA----- 6951  
QY 4948 QLGDOTCLLIAGLDRAAGCYTCQVSN-----KFGQVTHSA 4984  
Db 6952 -----YAEVLISPSERSDKGYTLKLENRYKTIISGEIDVNVVIAPSPAKELKFGDITKDS 7006  
QY 4985 CVVSGSEBAESSGGELDDAFRRAARLHLRFRK----- 5021  
Db 7007 VHLT-----WEPDDDDGSGPLTGYVVEKREVSRTKWTVMDFVTDLFTVPLVQKREYLF 7062  
QY 5022 -----SPAESDEELFLSADGCPAEPPEPADWQTYREDEHI----- 5058  
Db 7063 KVCARNKCGPEPAV-DEPVNMSTPATVDPDPENVKWRDRTANSIFLTDWPPNQGGSR 7121  
QY 5059 -----CIR-----FALTEARQAVTRFOE--MFATLIGIVBEIKLVEQPRRVM 5100  
Db 7122 IKGYIVERCPRGSKWACGEPVAETKVEVTGLEEGKWA-----YRKVT 7166  
QY 5101 CISKETAPVPPPEPLSLTSDRAAPFLATE-LQNOEVQDGYVPSFDCVVCQWPSVR 5158  
Db 7167 LNRGAKSPRPTETIQAQVDTQEAPEIFLDVKLAGLTVKAGTKIELPATVTKRPEKIT 7226  
QY 5159 WFKDGKLEEDHYMINEDQGGHQLIITAVPADMGVYRCLAEBSNGVSSKAEELRVDL 5218  
Db 7227 WTKADMILKQKRITI-ENVPKSTVTIVDSKRSDTGTIIIEAVNVCGRATAVVENV-L 7284  
QY 5219 TSTDYDTAADAESSYFSAQGYLSREQSGTSTTDEGQLPQVVEZLRDLQVAPGTRLA 5278  
Db 7285 DKPGPAAFDITDVNESCCLTWMPPRDDGGS----- 7316  
QY 5279 KFLKVGKYPAPRLYWFKDGQPLTASAHIRMTGKILHLEIISVTRDSGQVAAVISA 5338  
Db 7317 -----KITINYVER-----RATDSEVWHL-----SSTVKOINFKATKLIPN 7353  
QY 5339 MGAAYSSARLLVRGDEFEKPSADSVHEQLVPPRMLERFTPKVKKGS-SITFSVKVE-- 5395

Db 7354 KEYIFRVAENMYGAGEPVQASPIAKYQDFDPGPPTRLPESDITKDVAVTLTWCBPDDGG 7413  
QY 5396 GRPPTVHVLREBEAERGLWIGFDPGTGYTVASSAQOHSILVLLDVGRQHGTY--TCIASN 5453  
Db 7414 GSPI-TGYVW-----ER-----LDPDTDKVRCNKMPVKDITYRVKGLTNKKKYRFRVAEN 7464  
QY 5454 AAGQALCSASLHVSGLP-KVEQEKVKREALISTFLOG--TTOAISAQGLFETASPADLGGQ 5510  
Db 7465 LAG-----PGKPSKSTEPILIKPIDPWPFPKGTVDKVGKTSVR-----LAWT 7508  
QY 5511 RKEEPLAAKEALGHLSLAEVGTETEEFLQKLTSQITEMWSAKITQAKLQVPG-GDSDEDSKT 5569  
Db 7509 KEHDGGALESYVVIEMLKTGTDEWR-----VAEGVPTTQHLPLGLMEGEYSFR 7559  
QY 5570 PSASPRHGRSRSSSIQSSSESESDGARGEIFDIYVVTADYLPUGABQDAITIREGY- 5628  
Db 7560 VRAVNKAGESEPE-----PSDPVLCREKLYP 7586  
QY 5629 -----VEYLDAAH--PLRWLVRTKP-----TKSSPSROGW--VSPAYLDRRLK 5667  
Db 7587 PSPPRWLEVINITKNTADLKWTVPKDGSGSPITNIVIEKRDVRRKGWQTVDTTVKDTKCT 7646  
QY 5668 LSP--EWG-----AAEAPFPGEAVSEDEYKARLSSVIOELLSSQAFVEELOQSHH 5719  
Db 7647 VTPLTEGSLYPRVAAE-----NAIGQSDY-----TEIEDSLAKDTFT----- 7685  
QY 5720 LOHLERCHVPIAVAGXAVIFRNVDRDIGHSSFLQELQCCDITDDVAMCFI-KNOAAF 5778  
Db 7686 -----TGPPIYALA-----VVDTKRHVDLKWEPKNDGGRPIQRVVIKKEKELG 7730  
QY 5779 EQYLEBFLVGRV-----QAESVVVSTAIQEFYKYAE-----BAILAGDPSQ 5819  
Db 7731 TRWV--AGKTAGPCNFRVTDVIGTEVQ--FQVRAENEAGVGHSPSEPTILSIEDPTS 7786  
QY 5820 PPPPLQHYLE-----QOVER-----VORVOALLKELIRNKA 5851  
Db 7787 PPSPLDLHVTDAGRKHIAIAWKPEKNGSGSPITGYHVEMCFVGTETKMRVNSRPIKDLK 7846  
QY 5852 RNRQNCALLEQAYAVVSALPORAENKLHVS-----LMENYPG-----TLEALGEPIRQ 5900  
Db 7847 FKVESGVVPDKEY-----VLRVRAVNAIGVSPSEISENVAKDPCKPTID-----LETH 7897  
QY 5901 HPIVWEG-----APGARMPWKGNHRV-----FLPRN-----HLVICKP-RRDS 5938  
Db 7898 DIIVTEGEKLSIPVPFRAVPVPTVSMHKDGKEVKASDRLTKNDHISAHLEVPKSVRADA 7957  
QY 5939 RTDTSYVFRNMKLSIDLNDQVEG-----DDRAFEV-----WQ----- 5973  
Db 7958 GIYTT--LENKLSATASINVKIGLPGCKDIKASDITKSSCKLTWEPPEFGGTPIL 8015  
QY 5974 -----EREDSVRYLLQAFATAIKGSW-VKEIC-----GIQBRALPLV- 6010  
Db 8016 HVLBEREAGARTYIPVMSGENKLSWTVKDLIPNGEYFRVKAHVKNVGGGSEYIELKNPVI 8075  
QY 6011 ---WEPPD--FEELADCTAELGETVKLACRVGTGTPKFSVSWYKDGKAVQVDHHLIE 6064  
Db 8076 AQDPKQPPDPVDEHVNHTAE-----AMTITWKPLP--YDGSKIM--GYIEK 8121  
QY 6065 DFDGS-----CALILDSLTGVDSG---QYMCFAASAAGNCSTLGLKILVQVPP-RF 6110  
Db 8122 TAKGEERKRCNEHLVPLITVTAKLEEGKEYQPRVRAENAAAG---ISEPSRATPTPKA 8177  
QY 6111 VNKVPAASFV-----EGEDAQFTCTIECAPYQIRWYKD-----GALLTT 6150  
Db 8178 VDPIDAPKILRTSLEVKRGEIDALDASISGSPPTITWIKDENVIPEETIKGAAPLVR 8237  
QY 6151 GNKFO-----TLSEPRSGLLVIVIRAASKEDGLYECELVNRLGSARASA 6195  
Db 8238 RKGEVQEEEPFVPLTQRLSIDNSKKGESQLRVRDSILRPHGLYMLKVENDHGIAKAPC 8297  
QY 6196 ELRTQS-----PMLQAOECHOCHREQLVAVE-----DTTLERADQZ-----VT 6232

Db 8298 TVSVLDTGPPINFVFDIRKTSVLCKWEPPLDDGSEIIINYITLKKDKTKPDSEWIVVT 8357  
 Qy 6233 SVLK-----RLLOPK-----APGPP-----TGDLTGPGPCP----- 6258  
 Db 8358 STLRCHKSVTKLIEGKEYLFRVRAENRFGPPPCVSKPLVAKOPFGPPDPKPIVEDV 8417  
 Qy 6259-----RGAPAL----- 6264  
 Db 8418 TSNMLVKNPEKONGSPILGWLEKREVNTHSRVKNLSLNALKANVDGLLEGLTYVF 8477  
 Qy 6265---QSTGQPPVTGSEAPVPRVPOPLLHGGPEQPEAJARAQEWTPVIRMEGAAPWG 6321  
 Db 8478 RYCAENAAAGP---KGFSPSPDKTAHDPISPPG---PIPRVTDTSSTTIELEWPPAFNG 8532  
 Qy 6322 AG-----TCELLWD-----VSHVIVRETTQ-RTYTVOAIDTITA----- 6354  
 Db 8533 GGEIVGYFVDKOLVGNKWSRCTEKVIMKVRQYTVKEIREGADYKLRVSAVNAAGEGPPE 8592  
 Qy 6355-----RPPSMQVITIE---DVQAOTGGTAQPEAIIIEGDDPOPSVTVYKDSVQLVDST 6401  
 Db 8593 TQPTVVAEPQEPVAVELDVSKGGIIMAGKTLRIPAVVTGTPVTKVTKKEGEL-DKD 8651  
 Qy 6402 RLSQOQEGTYSVLVRHVASKDAGVYTCLAQNTGGQVLCKAELLVLGGDNEP-----DSEK 6457  
 Db 8652 RVVIDNVGTSKSLIIKDALRKDHGRYVITATNSCGSKFAAARVEVF---DVPGPVLDLKP 8708  
 Qy 6458 QSHRHK---LHSPVEKKEIGRGVFGVKVQKHGKILCAAKFIPLRSRTAQAVRREDI 6515  
 Db 8709 VVTRKMLLNSDPEDDGGSEITGFI---IERKDKM----- 8743  
 Qy 6516 LAALSHPLVTGLDOPETRKTLILILELCSSELDRLRYKGVVTEARVKYVIOOLVEGL 6575  
 Db 8744-----RGAPAL----- 6766  
 Qy 6576 HYLHSHGVL-----HLDIKPSNIMVHAREDIKIDGFPAGNITPAELQFSQVGS 6626  
 Db 8767 EV---KFRVAKNKGFGPPVEIGP---ILAVDP-----LGP 8797  
 Qy 6627 PBEVSPETITQONPVSEASDIWANGVISVLSITCSPFFAGESDRATL-LNVLEGRVSWSSP 6685  
 Db 8798 P---TSPE-----RLTYT-----ERQSTITLDWKEPRNGGSP 8828  
 Qy 6686 MAHULSE---DAKPIKATLORAPQAPSAQCLSHPWFLKSMFAEAEAHNTKQLXEL 6741  
 Db 8829 IQGYIIEKRHRDKPDF-----ERNVKLCPTTFLVENLDEHQMYSFRVK----- 8873  
 Qy 6742 LARSQRSLMSYKSLVMRSIPELLRGPDPSPSLGVARHLCDRTGGSSSSSSSDNELA 6801  
 Db 8874-----AVNEIGSEPSLP-LNVV-----IQDDEVP 8897  
 Qy 6802 PFARAK-----SLP---PSPVTHSLP-----LHPRGFLRPSASL---PEAE 6837  
 Db 8898 PTIKRLSVRGDTIKVKAGEPVHIAADVTLGLPMPKIEMSKNETVIEKPTDALQITKEVS 8957  
 Qy 6838 ASERSTEAPAPASPEGAGPPAAQGVPRHSVIRSLFYHQAGESPEHGALAPGSRHHPAR 6897  
 Db 8958 RSEAKTELSIPKAVREDKTYVTVNASNLGSRVFRNVHVEYDR-----PSPPRNLAV 9009  
 Qy 6898 RRHLLKGGYIAGALP---GLREPLMEHVRVLEEAAREEQ-----TLLAK-----A 6940  
 Db 9010 TDIIKAEBCYLTDWAPLDNGSE---ITHYVIDKRDSARKKAEMEEVNTAVKRYGIWKLI 9067  
 Qy 6941 PSFETALRIPASGTHLAPGSHSLHSDSPSTRPSSACEAQRLPSPSGAPIEDMGH 7000  
 Db 9068 PNGQYFRVRANKY---GISDECKSDKVIQDP-----YRLPGPP---GKP-KVLAR 9113  
 Qy 7001 PQGSKOLPST-----GGHPGTAQ---PERPSPDSP-WGQAPFCHPKQG-----SAQP--E 7045  
 Db 9114 TKGSMVSVTPPLDNGSGSITGWLEKREGSPYMSRVSRAPITKVLKGVEFNVRLLLE 9173  
 Qy 7046 GCS-PHPVAPCPGFPFGSGCKEAPLVPSSP-FLGQOAPAPAKASPLDLSKMGPGDI 7103  
 Db 9174 GVKYQFRAMAINAAGIPSPSE-----PSDPEVAGDPIFPFGP-PSCEVVDKT-KSSI 9224

Qy 7104 SLPRKFPGPCSPFGSQAQSSQVSSLRVGSQVCTEPGPSILDAEGW---TQAEPLSDS 7161  
 Db 9225 SLGWKP---PAKGG---SPIKGYIVEMQEGT-----TDWKRKVNPEPKLIIT 9266  
 Qy 7162---TPTLQRPQOATMRKFSGLGGRGYAGVAGYGTFAFGDAGGMLGQGPMMARIJAWAV 7217  
 Db 9267 CECVVPNLKE-----LRKYRF-----RVK-AV 9287  
 Qy 7218 SQSEEEBQEARAESQSEBQEARAESPLPOVSARPVPEVG-----RAPTRS 7264  
 Db 9288 NEAGESEPSDTTGEIPATDIQES-----PEVFI---DIGAQCULVCKAGSIOIRPAVI 9337  
 Qy 7265 SPPEPTP---WEDIGQVSLVQ---IRDLSDGAEEAANDTISLISEVDPAAYLNLSLDYDI--K 7316  
 Db 9338 KGPTPKSSWEFGKAKKAMKDGVDHIDPEDAQLETAENSVIIIPCKRSHTKYSITAK 9397  
 Qy 7317 YLPEFPMIFKPKSAQPPSPMAEBELAEFP-EPTWPPGELG----- 7360  
 Db 9398 NKAGQKTANCRCVKMVDVPGPPKDLKVSIDITRGSCRLSKWMPDDDDGDRINKGYVIEKRTID 9457  
 Qy 7361-----PHAGL-----EITESEEDVDALLAAAVGRKRKWSPPSRSLFHPGGRHLPL 7406  
 Db 9458 GKAWTKVNPDCGTTTFVVDLLSEQQYFFRVRAENRFGIGPPVETIQTARTDP--LYPP 9515  
 Qy 7407 DEFAELGLRERVRKASV-----EHSIRILKGRPEGLEKEG----- 7440  
 Db 9516 DPKIKLIGLITKNTVHLSWKPKNDGSPVTHIVECLAWDPTGTKEAWROCKNRDVE 9575  
 Qy 7441-----PPRKKPGLASRSLGLKSW-----DRAPTF-LRELSDETIVL 7476  
 Db 9576 ELQFTVEDLVEGGEYFRVAVNAAGVKSPTSATVGCDCORPMDPSPIDLUKEME--VEE 9633  
 Qy 7477 GQSVTLACQVSAQAPAAQATWSKDGAPLESRRSLISATLKNFOLL--TILVVVAE---D 7530  
 Db 9634 GTNVNIVAKIKGVPPFTLTWFKAPPKPDNKEPVLVDTHVKNLWDDTCTLVIPQSRSD 9693  
 Qy 7531 LGVITCSVALGTVTTTGVLRKAERSSSPCPDIDG-----EYVADGVLLVWKEVSYGP 7585  
 Db 9694 TGLYITITAVNLGTASKEMRLNVLGRPG---PVGPIKESVADQMTLSWFPKDDGG 9749  
 Qy 7586---VTVYIWO---CSLEGGSWTTLASDIPCCYLTGKLSRGYTYTERTACVSKAGMG-PYSSP 7640  
 Db 9750 SKITNYVIERKRNKTVHVSSEPEKCTYIPKLEHGEYVFRIMAKNYGIGEPDSE 9809  
 Qy 7641 SE 7642  
 Db 9810 PE 9811

RESULT 9

Q8WZ42 ID Q8WZ42 PRELIMINARY; PRT; 34350 AA.  
 AC Q8WZ42;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE titin.  
 GN TTN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20309627; PubMed=10850961;  
 RA Freiburg A., Trombitas K., Hell W., Cazorla O., Fougereuse F.,  
 RA Centner T., Kolmerer B., Witt C., Beckmann J.S., Gregorio C.C.,  
 RA Granzier H., Labeler S.;  
 RT "Series of exon-skipping events in the elastic spring region of titin  
 RL as the structural basis for myofibrillar elastic diversity";  
 RL Circ. Res. 96:1114-1121(2000).  
 RN [2]



Db 4395 ALGHLAKFCEIQSAPNVRWFOWFKAGREIYESDKSIRSSKYISSILEILRTQVDCGEVT 4454  
QY 1054 CEARGQVRF---RLHITE---PKMFAKEQSVHNEVQAEAGASAMLSCEVAQAQT-EVT 1106  
Db 4455 CKASNEYGSVCTAILTVTEAAPPFLSRPKSLTTFV---CKAAKFICTVGTPIVETI 4510  
QY 1107 WYKDGKLSSSKVGNEVKGCTRRLLVPOAGKADAGEYSCEA---GGQVRSFHLHITEP 1162  
Db 4511 WOKDGAALSPSPNWKISDAENKHIELSNLTQDRGVYSCKASNFGADICQAEILIID- 4569  
QY 1163 KGVFAKEQSVHNEVQAEAGTTAMLSCVAQP-QTEVTWYKDGKLL--SSSSKVRNEVKG 1219  
Db 4570 KPHFIKEL---EPVQSAINKVHLBCQVDEDRKVTWWSKDGCKLPPOKDYKICEDKIA 4626  
QY 1220 TRRLVVOQKADAGEYSC---EAGGORVSPQLHITEPKAVFAKEQVLVHNEVRTEAGAS 1275  
Db 4627 TLEIPLAKL--KDSGTIVCTASNEAGSSCSATV-VREPPS-FVKK--VDP SylMLPGES 4681  
QY 1276 ATLSCEVAQAQT-EVTWYKDGKLSSSKVRTEAAGCWRQLVVOAGQADAGEYTCEA- 1332  
Db 4682 AELHCKLKGSPVQVTFWFKNKLSESNTVMFYVNSEAILDITDKVEDSGSYSCAEVN 4741  
QY 1333 --GGQRLSFHLDOVBPKAVFAKEQLAHRKVQAEAGAIATLSCEVA-QAOTEVTWYKDGK 1389  
Db 4742 DVGSDSCSTEIVIKPPSFIKLEPA---DIVRGTNALLQCEVSGTGPFIISFKDKKQ 4797  
QY 1390 LSSSKVRNEAVGCTRRLLVVOAQCOADTGEYSC---EAGGORLSFLDVAEPKVVFAKE 1445  
Db 4798 IRSSKRYLFSQKSLVLEIFSNFADYGEYCVVANEVKGCGCMATHLLKEPPTFVKV 4857  
QY 1446 QPVHEVQQAQAGASTLSCEVAQAQ-TEVMYK-----DGK-KLSFSSKVRNEAVGCT 1496  
Db 4858 D-----DLIALGQVTLQAVRGSEPI SVTMKGOEVIREDGKIKMSFNGAV----- 4907  
QY 1497 RRLVVOQAGADAGEYSC---EAGQRLSFHLHVAEPKAVFAKEQBPASREVQAEAGTSA 1552  
Db 4908 --LIIPDVQISFGKYVCLAEAGSQTSGELIVKEPAKIER---AELIQVTAGDPA 4961  
QY 1553 TILSCVA-QAOTEVTWYKDGKLSSSKVRNEAVGCTRRLLVVOEAGQADAGEY-----SK 1607  
Db 4962 TLEYVAGTPELPKWKYKDGRLVASKYRISFKNNVAQLFYSAELHDSQYTFEISNE 5021  
QY 1608 AGDQRLSFHLHVAEPKAVFAKEQPAHREVOEAGASATLSCEVA-QAOTEVTWYKDGK 1666  
Db 5022 VGSSSCETFTVLRDRIAPFTTKPL-RNVDSVVGTCRLDCKIAGSLPMRYSWFKDGKEI 5080  
QY 1667 SSSKVRNEAVGCTRRLLVVOQAGADAGEYSCA---GGQRLSPLHVAELFQISERP 1722  
Db 5081 AASDRYIAFVEGTASLEIRVDMDNAGNFTCRATNSVGSKDSSGALIVQES-PPSFVTKP 5139  
QY 1723 CREPLVKEHBDIILTATLTPSAATVTLKDGVEIRRSKRHETAPASQGDTHTLVHGAQ 1782  
Db 5140 GSKD---VLPGSVCLKTFQGSTPLTRNFKGNKEIVSGGSCYITKEALESSELYLVK 5196  
QY 1783 VLDSAIYGRVG--AEGQDPVQ--VEEVAAKFCLLLEPPVCGEL---GGTVTLACELS-P 1834  
Db 5197 TSDSGTYTCVSNVAGGVECSANLFVKE-PATFVEKLEP--SOLLKKGDATOLACKVGTGT 5253  
QY 1835 ACAEVVRCGNTQPVKGFQM--VAEGPVRSLTVLGLRAEDAGEYCESRD---DHTS 1888  
Db 5254 PIKITWANDREIKESSKHRMSFVESAVILRLTDVGI--EDSGEYMCQAEAGSDHCS 5311  
QY 1889 AQLTV-SVPRVVKFMSGIST-----VVAEGGEATFCQVVSFSDVAVVMFRGALIQP 1940  
Db 5312 SIVIVKESPYTFKEFKPIEVLKEYDVMLLAEVAGTPPE-----ITWFKONTILRS 5362  
QY 1941 SKFAISQSGASHSTISDLVLEDAGO-----ITVEAGASSAALVRBEAPVLFKKLEP 1996  
Db 5363 GRKYTFIQDHLVLSQILKFVAADAGEYQCRVTNEVGSSICSAVTLREPP-SPIKKIES 5421  
QY 1997 QVVEERSVTLVELTRWEP-ELRWTRNATALAPGKNVEIHAEAGRHRLVLHNVGFADRG 2055

Db 5422 TSSLRGGTAAQATKGLSLPITVTWLKDSDEITDDNIRMTFENNVAASYLSLGVIEVKHDG 5481  
QY 2056 PFGCETEDD---KTOAKLTVEMRQVRLVRLGLQAVEAREOGTATMEVOLSHA-DVDSGWT 2110  
Db 5482 KYVQAKNDAGIORCSALLSVK-EPATITEEAVSIDVTQGPATLPQVPSGTKEITAKWF 5540  
QY 2111 RDGLRFQOGPTCHLAVRGPMTHTLTLGSLRPEDSGLMVFFKAEG---VHTSARLWVTEL-- 2164  
Db 5541 KDGQELTLGSKYKISVTDVTSILKIISTEKKDSGEVTFEVDNVRSSCKARINVLDLII 5600  
QY 2165 PVFSRPLQDVVITEKEKVTLECEL--SRPNVDVRLKDGVELRAGKTMAIAAQAQACRS 2222  
Db 5601 PPSFTKLLKNDSTIKGSFIDLECIVAGSHP-ISIOWPKDQOEISASEKEYKFSFHDNTAFL 5659  
QY 2223 TIYRCEPADQGVVYVCAHDAQSSASVKVQGRTYTLIYRRLAEDAGEIQVAENASRAQ 2282  
Db 5660 EISLEGTDGSGTYCSATNK-----ACHNQCSGH 5688  
QY 2283 LRVKELPVTILVRPLRDKIAMEKIRGULECOVSRASQVRFKSQBLQOPGKVELVSDGL 2342  
Db 5689 LTVKEPPYFVEKPSQSDVNPENTRVLKALVGGTAPMTIKWFKONKELHLSGAARSVMKDDT 5748  
QY 2343 YRKLIIISDVAEDEDVTVC---DAGDVKTSAOFFVEE-----QSIITIVGLQDVTV 2390  
Db 5749 STLSLELPAKATDSGTIYICQLSNDVGTATSKATLVKEPPQPIKPSVLVLNAGSQT- 5807  
QY 2391 MEPAPAWFECE-TSIPSVRPPKMLGK--TVLOAGG-----NVGLEQEGTVH 2434  
Db 5808 -----FECQITGTPKIRVSWYLDGNEITAIKOHGISFIDGLATFQISGARVNSGT- 5859  
QY 2435 RLMLRRCTMTGPHFTVCGKSSARLVSDIPVVLTRPLEPKTGRELSVVLSCDFRP 2494  
Db 5860 -----VCEARN-----DAGTASCSTELKVEPPTI-RELKPEVVVKYSDVLECEVTG 5907  
QY 2495 APK-AVOMYKDDTFLSPSEKFMKLEGQMAELRLMLPADAGVYRC-----QAGSAHSST 2549  
Db 5908 TPFPEVTWLNKONREIRSKKYTLTDRVSVFNHLITKCDPSDTCGEYOCIVSNEGSCSCT 5967  
QY 2550 EVTVE-----AREVTVTGPLQDABATEEGWASFCLSHEDEREVEHSLNGMPLYNFSFHE 2604  
Db 5968 RVALKPSPFIKIENTTTVLKSSATFQSTVAGSPFIS-----ITWKDQDILDEDDNVY 6022  
QY 2605 ISHKGRRHTLVLXSIQRADAGIVRASSLKVSTSRARLEVRV-----KPVVFLKALD 2654  
Db 6023 ISFVDSVATLIQIRSVNDHSG-----RYTQAKNESGVERCYAFLVQEPQAVIVEKAK 6075  
QY 2655 DLBAERGTLALOCYS-DPEAHVVRKQVQLGPSQKYDFLHTAGTGRGLVHVDVSPEDA 2713  
Db 6076 SVDVTEKDPMTLECVVAGTPELKVKWLKDGKQIVPSRYFSMSFENNVAIFQSVNMKQDS 6135  
QY 2714 GLYTCHV---GSEETRAKVRVHDLHV--GITKRLKTMVELEGESCSFECVLSHESASDP 2767  
Db 6136 GQYTFKVENDFGSSCDAYLRVLDQNPSPFTKLTMDKVLGSSIHMECKYS-GSLPIS 6194  
QY 2768 AMWTVGKTVGSSSRFQATROGRKYILVVREAAPSDAGEVVFVSRGLTS-----KASLIYR 2823  
Db 6195 AQWFKDGKEISTSAKRYLVCHERSVSLVNNLEEDTANYTCKVSNVAGDACCGLTVK 6254  
QY 2824 ERPAALIKPLEDOWAPGEDVELRCELRAAGTP---VHMLKDKKAIKRSQKYDVVCEGTM 2880  
Db 6255 EPSFVLVFGROQAI-PDSTVEFKALK--GTPPFKIKWFKODVELVSGPKFIGLEGST 6311  
QY 2881 AMLVIRGASLKDAGETYC---EVEASKSTASHVEEKANCETELTNLQVEBK--TAYF 2935  
Db 6312 SPLNLVSDASKGTQYTCVHTNDVGSCTTMLLVTEPPK-FVKLEASKIVKAGDSRL 6370  
QY 2936 TCK-TEHPAATVWRKGLLELRASGHQPSQEGTLRLTILSALEKADSDTYTCD-----IG 2990  
Db 6371 ECKIAGSPEIRVWVFNHEHLPASDKRYMTFIDSVAVIQMNILSTEDSGDFICEAQNPA 6430  
QY 2991 QAQSRQALLVQGRVH-----IIEDLEDVDVQEGGSATPRCIRISPAHYFVHFWLKTPL 3045  
Db 6431 STCSSTKIVKPEPVSPPPIVETLKNAEV-----SLECELSGTPPEVWYKDKROL 6484







Db	8482	VEFANAGKTCQIKNDAGMOECFAILSVLEPATIVK-----PSI	8522
Qy	5117	PSLLTSDAAPVLTQLQNOEQDQGVPSFDCVVVTCQPMPSVRFKDGKLLBEDDHYMIN-	5175
Db	8523	-----KVTGDTCTLECTVAGTPPELSTKWFKOGKELTSDNKYKISF	8563
Qy	5176	EDQOQGHQIITAVVPADMGVYRCLAESMGVSSYKAEKRLDVLTDYDVTDAADATESSY	5235
Db	8564	FNKVSQ--LKIINVAPSDSGVYSFEVQNPVVGKDSCTASLOV-----	8602
Qy	5236	PSAQYLSRREOEGESTTDEQQLFQVWEELRDQVAPQTRLAKPQKVGYPAPRLYWF	5295
Db	8603	-----SDRTVPSPFTRKLTNGELSGSSVV-NECKVYSGPPIVSVMF	8643
Qy	5296	KDQOPLTASAHIMTKKILHLLEIISVTRSDSGQYAAVISNAGAAYSARLLVGRPDE	5355
Db	8644	HEGNEISGRKYOTLTNDNTCALTNMLEESDSGYTCIATNMAGSDSCSAPLTVR----	8699
Qy	5356	PEEKPADVHEQOLVPPRMLETRPKVKKGSGIITPSVKVGRPVPTVHMLREAEARGVLW	5415
Db	8700	-----EPPSFVQKDPMDVLTGTNTFTSIVKGTTPPSVSMFKGSSBL-----	8742
Qy	5416	IGPDTFG--YTVASSAQHSLVLLDVGRHOHQTYTCIASNAGQALCSASLHVSGLPKVE	5473
Db	8743	-----VPGDRCNVSLDSVAELELFDVDTSQSEYTCIVSNEAGKASCTHLYIKAPAKFV	8798
Qy	5474	EQ-----EKKVEALISTELOT--TOAISA-----QGLETASPADLGGORKEEPLAAKE	5520
Db	8799	KZLNDYSIEKGXPLI-----LEGTFTGTPPISVTKNGKGINVTP-----SORCNITTEKS	8849
Qy	5521	ALGHLSLAEVGTTE-----FLQKLTQSQITEMVSAKITQAK-----LQVPGGSD	5564
Db	8850	AI--LEIPSSSTVEDAGQVNCYENASGK--DSCSAQILILEPPYVKQLEPVKVSVDGSA	8905
Qy	5565	EDSKTPSAPRGRS-----RPSSS-----IQES-----SSSE	5593
Db	8906	SIQOQLAGPEIGVSWYKGDTKLRPTTYKMFNRNNTLVFNQVDINDSGEYICKAENS	8965
Qy	5594	DGDARGEIPDIYVVTADYPLGAEQDAITLREGQYVEV-----LDAAHPL--RMLVTRKP	5646
Db	8966	VGEVSASTF--LTVQEQKLPPSFSRQLRDVQETGLPVVFDCAISGSEPIVSVMYKDGK	9023
Qy	5647	TKSSFSRQGWSPAYLDRLKLSPEWGAAPPEPFGAEVSEDEYKARLSSVIOELLSEQ	5706
Db	9024	LKDSFN--VQTSFLDNTATLN-----IFKTRSLAGQVSCATNPIGSASSAR	9070
Qy	5707	AFVEELQFLQSHHLOHLERCPHVPIAVAGQKAVI-----PRN	5743
Db	9071	LILTE-----GKNPPFFDRLAPDAVVGESADFECHVTGTQPIKVSNAKDSRE	9119
Qy	5744	VRDIGRFHSSFLQ-----LOQCDTDDVAMCFIKNQAAFQYLEFLVGRVQAESVVVST	5798
Db	9120	IRSGGKIQISYLENSAHLTVLKVKGDS-----GQYTCYAVNEVGKDSCTAQL	9167
Qy	5799	AIQEFYKKAEBALLAGDPQPPPLQHYLEQPVYRQVQALKEILIRKARNQONCA	5858
Db	9168	NIKE-----RLIPSFYKRLSEET-----	9189
Qy	5859	LLEQAYAVSALPQRAENKHLVSLMENYPTGLEALGEPIROQHFTVWEGAPCARMP----	5914
Db	9190	-----EGNSFKLEGRVAGSQPIVA	9209
Qy	5915	WKGHNRHV-----FLFRNLHVICKPRDSRTDVTVSIVFRNMKLSLIDNDQVEGDDR	5967
Db	9210	WYXNNIEIOPTNCNCEITFKNTLVIVQRKAGMNDAGLYT-----CKVS-----ND	9254
Qy	5968	AFEVQWEREDSVRYLLQARTAIKSSWV--KEICIGIQORLALPVNRPPDFEELADCTAE	6026
Db	9255	-----AGSALCTSSIVIKE-----PKKPPVFDQHLTPVTVS	9285
Qy	6027	LGETVKLACRVGTGPKPVISWYKQKAVQVDPHILIEDPGSCALILDSITGV-----	6080
Db		-----	
Db	9286	EGEVVOLSVCHVQSEPIRIQWLKAGREIK-----PSDRCSFSFASGTAVLELRDV	9335
Qy	6081	---DSGQYMCFAASAAGNCSTLQKILVQVPPR-----FVNKVRASPPFVEG	6122
Db	9336	AKADSGDYVCKASNVAGSDTKSKVTIKDKPAVAPATKAAVDGRLPFFVSEPQSIRVVEK	9395
Qy	6123	EDAOFTCTIEGAPVQIRWYKDGALLTTGNKFTLSB-----PRSGLLVLVIRASKE	6175
Db	9396	TTATPIAKVGDDIPNVK-----TKGKWLQNOGGRVFIHQKDEAKLEIRDTTKT	9447
Qy	6176	DLGLYECELNRLGSRASAEALRI-----QSPMLQ-----	6205
Db	9448	DSGLYRCVAFNEHGETESNVNLQVDERKKQKEIGDLRAMLKKTPIILKKGABEEEDIM	9507
Qy	6206	-----AOEQCHREQLVAAVEDITLERADQ	6229
Db	9508	ELLKNVDPKYEKYARMYGITDFRGLQAFELLKQSOEETHRLEIEEERSERDEKEFE	9567
Qy	6230	EVTSLVKRLGPKAPGPGSTGDLTGPGCP-----RGAPALQ--	6265
Db	9568	ELVSFIQORLSQTEPVTLIKDIENQTVLKNDVAFEDIKINYPEIKLSWYKTEKLEPS	9627
Qy	6266	-----ETGQPPVTG-----TSAPAPVPVPPOPLHHEGP	6295
Db	9628	DKFEISIDGRHTRLVRKNCOLKQGNRYLCVCGPHIASAKLTVIEPAWERHLQDVTLKEG-	9686
Qy	6296	EQSPALARAQEWTP--IRMEGAAW-----PGAGTCELLDWDVHSV--VRETTOR	6342
Db	9687	-----QTCMTQCQSVNVKSE---WRNGRILKPOGRHTEVEHKVHKLTIADVRAEDQ	9739
Qy	6343	TYT--YQAITDHT-----ARPPSMQVTEIDVQAGTGAQFAIIEBGPQPSVTVYKDSV	6395
Db	9740	QYTCYKEDLETSAELRIEAPQIOTKRIQINIVSEHQSAFCEVEVSFD-DAIVTWYKGP	9798
Qy	6396	QLVDSFELSQQGTTYSVLVRHVASKDAGVYTCQAQ--NTGGQVLCQAEI-----LV	6446
Db	9799	ELTESQYNPRNDRCHYMTIHNVTDDGCVYVIARLEPRGEARSTAEYLITTKEIKLE	9858
Qy	6447	LGSDNEPDS-----BKQSHRRKLHSEFYVEKE	6473
Db	9859	LKPPIDPSRVPIPTWPIRAVPPPEIPVAVPIPLLPTPEEKXPPKRI-----EVTK-	9913
Qy	6474	IGRVFGFVRVQHKGNKILCAAKFPLRSRTAQAYRERDILAAASHPLVTLGLOFET	6533
Db	9914	-----KAVKDKAKVAKFKEMTPREEIVKXPPPTTILIPAKAPEII-----DVSS	9959
Qy	6534	RKTLILILELCSSEEL--LDRLYKGVWTEABVKVYIQQOLVEGLHLYHSHGVHLHDIKP	6590
Db	9960	KAEVKIMTITRKKEVQKEAEVYKQAVHKEKRVFIESFEFPD-----ELEVEP	10011
Qy	6591	SNILMVHP-----AREDIKI-----CDFGFAQNIPTAELQFSQYGSPE----	6628
Db	10012	YTBPFQPYEEDEDEYEEKVAKKEVHEWEDFEQGEYERBEGYDE-GEEMWEEA	10070
Qy	6629	FVSPETIQ--QNPVSEAS-----DIWAMGVISYLSLTCSSPFFAGESDR	6669
Db	10071	YQREVIQVQKEVYESHERKVPKAPPEKAPPPKVPKVPKVPKVPKVPKVPKVPKVPKVP	10129
Qy	6670	ATILNLVGRVSSWSPMAAHLSEDAKDFIKATLQAPQAPPSAAQCLSHFWFLKSPAAEE	6729
Db	10130	VTKVPEVSKKIVPKQSRTPVQSEVTEV-----KVPVAVTKKWVISEEKGFASFHTEEE	10183
Qy	6730	AHINTKQLKFLARSRWQSLMSYKSLVMRSIPPELLRGPPDSPLGVVARHLCRTGGS	6789
Db	10184	VS-VTVPEV-----QKEIVTEBKIHV--AVSKRVPPPPKVPKEL-----	10218
Qy	6790	SSSSSSSDNELAPFAKASL--PPSPVTHSPLLHPRGFLRPSASLPEAEASERSTAP-A	6847
Db	10219	--PEKAPAEVAVPPIPKKVEPAPKVPV-----PKKVPPEKKVPVPVPPKPEPAA	10267
Qy	6848	PPASPE-----GAGPPAAQCQVPRHSVI--RSUFYHQAGSPSPHGA-	6886
Db	10268	PPKVPPEKKVPPEKIPVFAKKEAPPAKVPQVQKRVVTEBKITVITVQREESPPPAVP	10327

QY 6887 -----LAPGSRHPARRHLLKGGYIAGALPGLREPLMEHRV-----LE 6925  
 DB 10328 EIPKKVPEERKPVPRKEEYPPPPKVPALP--KPVPEEKVAVPVVAKAPPPRAEVS 10385  
 QY 6926 EBAAREEATLLAKAPSETARLPASGTHLAPGH--SHSLEHDSPS-----6970  
 DB 10386 KKTVEEKRFVAEEKLSFAVQCRVETRHEVSAEENSYSEEGVSIIVREBEREEB 10445  
 QY 6971 -----TRPSSEACGE-----AQLPSAPSGGAPIRDMGHPQSGK-----QLPSTG 7011  
 DB 10446 EAEVTEYEMBEPEYVVEEKLHIISKVVEAEP-----EVTREQKVIKPKPAKI 10499  
 QY 7012 GHGTAQ--PERSPDSNGO--PARCPKOGSAQEGCSPHAPAVAPCPGSPFPFGCK-E 7068  
 DB 10500 BEPPPAKVPEAPKTVPEKKVPAPV--PKKEKVPKVPPEPK-KPVPEKKVPKVKME 10556  
 QY 7069 APL-----VPSSPFFZGQOAPPAPAXAGPPLDSKMG--PDISLPGRPKPGPCSS 7116  
 DB 10557 EPLPAKVTERRHQITQEBKVLVAVTKKAPPKARVPSEPKAVPEEKVLLKPKRER-EP 10615  
 QY 7117 PCSASQASSQVSSLRVGSSQVGTPEPGLDAEGWTQE--AEDLSDSPTTLORPQOATM 7174  
 DB 10616 PAKVTEFRKRVVKEEKVIEAPKREPQPIKEVTIMEEKERAYTLEEBAVSVQREEEY 10675  
 QY 7175 RKFSLGGRGGVAGVAGYGTFAFGDAGGMLGQGPMMARIAVAVSQSEEEQEEARBSQS 7234  
 DB 10676 EYDYKEFEYEPTEEYDQY-----EYEEEREYERYEHE 10710  
 QY 7235 EQQARAEPLQVQASRPVEVGRAP-----TRSSPE---PTWEDIGQVSLVQIRDLS 7286  
 DB 10711 EYITEPEKPIPVKVPPEBPVPTKPKAPPKAKVLKXAVPEEKVPV-----10754  
 QY 7287 GDAAEADTISLDSVDPAYNLSLDYDKYLPFEFMIFRKYKPSAQPPEPPSPMAEBELA 7346  
 DB 10755 -----IPKLLKPPPK-VPEBPK 10772  
 QY 7347 EPEPTWMPGELPHAGLEITE-BSEDV-----DALLAAVAGRKRWSSPSR 7394  
 DB 10773 VPEE-----KIRISITKKEQVTEBPAKVPKRVVAAEKKVPVPRKEVAP-- 10819  
 QY 7395 SLFHPGR--HLPLD-EPAELGLRERVKASVEHISRII-----7429  
 DB 10820 -----PVRPEVPEKLEPEEVAEEV---VTHVEEYLVEEEBEEYIHEEEFITEBEVVP 10871  
 QY 7430 -----KGRPEGLEKEGPPRKKP 7446  
 DB 10872 VIPKVPVEPRKVPPEKKVPVVPKKEAPPKVP 10906

## RESULT 10

Q8NH4 PRELIMINARY; PRT; 658 AA.  
 ID Q8NH4;  
 AC Q8NH4;  
 DT 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Obscurin (Fragment).  
 GN OBSCN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Young P., Ehler E., Gautel M.;  
 RT "Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere  
 RT assembly";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ314904; CAC85751.1;  
 DR InterPro; IPR003599; IG-  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003598; IG\_C2.

DR Pfam: PF00047; ig; 4.  
 DR SMART; SM00409; IG; 6.  
 DR SMART; SM00406; IGC2; 3.  
 DR PROSITE; PS00835; IG\_LIKE; 4.  
 KW Immunoglobulin domain.  
 FT NON\_TER 1  
 FT NON\_TER 658  
 SQ SEQUENCE 658 AA; 72830 MW; CB37B453FCD4AE4A CRC64;  
 Query Match 8.2%; Score 3389; DB 4; Length 658;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-173;  
 Matches 658; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2075 ROVLVRGLQAVEAREEQGTATMEVQLSHADVDGSTRDGLRFQOQPTCHLAVRGPMHTLT 2134  
 DB 1 ROVLVRGLQAVEAREEQGTATMEVQLSHADVDGSTRDGLRFQOQPTCHLAVRGPMHTLT 60  
 QY 2135 LSGLRPEDSGLMVFKAGVHTSARLVVITELPVFSRPLQDVVTEKEKVTLECELSRNV 2194  
 DB 61 LSGLRPEDSGLMVFKAGVHTSARLVVITELPVFSRPLQDVVTEKEKVTLECELSRNV 120  
 QY 2195 DYRWLKDGVLELRAGKTMAIAAQAACRSITIRCEFAADQGVVCDADHAQSSASVKVQGT 2254  
 DB 121 DYRWLKDGVLELRAGKTMAIAAQAACRSITIRCEFAADQGVVCDADHAQSSASVKVQGT 180  
 QY 2255 YTLIYRVLAEADAGIQFVAENAESRAQLRVKELPVTILVRPLRDKIAMKRGVLECOVS 2314  
 DB 181 YTLIYRVLAEADAGIQFVAENAESRAQLRVKELPVTILVRPLRDKIAMKRGVLECOVS 240  
 QY 2315 RASAQVRWFKGQELQPGKPYELVSDGLYRKLIISDVHAEDEDTTCDAQDVKTSAQFFV 2374  
 DB 241 RASAQVRWFKGQELQPGKPYELVSDGLYRKLIISDVHAEDEDTTCDAQDVKTSAQFFV 300  
 QY 2375 EQQSITIVRGLQDVVMEPAPAFECETSIERSVPPKLLGKTVLQAGNVGLEQGTGH 2434  
 DB 301 EQQSITIVRGLQDVVMEPAPAFECETSIERSVPPKLLGKTVLQAGNVGLEQGTGH 360  
 QY 2435 RLMRLRTCTMTGPHFTVGRSSARLVSDIPVILTRPLEPKTGRQLQSVLSCDERP 2494  
 DB 361 RLMRLRTCTMTGPHFTVGRSSARLVSDIPVILTRPLEPKTGRQLQSVLSCDERP 420  
 QY 2495 APKAVQWYKDDTPLSPSEKFKMSLEGQMAELRILMPADAGVYRCQAGSAHSSTEVTE 2554  
 DB 421 APKAVQWYKDDTPLSPSEKFKMSLEGQMAELRILMPADAGVYRCQAGSAHSSTEVTE 480  
 QY 2555 AREVTVTGPLQDAEATEEGWASFCELSHEDDEWENGLMPLYNDSFHEISHKGRHTL 2614  
 DB 481 AREVTVTGPLQDAEATEEGWASFCELSHEDDEWENGLMPLYNDSFHEISHKGRHTL 540  
 QY 2615 VLKSIQRADAGIVRASSLKVSTASRLEVRKVPVFLKALDLSAERGTLALQCEVSDPE 2674  
 DB 541 VLKSIQRADAGIVRASSLKVSTASRLEVRKVPVFLKALDLSAERGTLALQCEVSDPE 600  
 QY 2675 AHVWRKQGVQLGSPDKYDFLHTAGTRGLVHVDVSPDAGLYTCHVGSEETRARVRVH 2732  
 DB 601 AHVWRKQGVQLGSPDKYDFLHTAGTRGLVHVDVSPDAGLYTCHVGSEETRARVRVH 658  
 RESULT 11  
 ID Q10465 PRELIMINARY; PRT; 7962 AA.  
 AC Q10465;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Elastic titin (Fragment).  
 GN TITIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

TISSUE=skeletal muscle;  
MEDLINE=96026330; PubMed=7569978;  
Label S., Kolmerer B.;  
"Titins, giant proteins in charge of muscle ultrastructure and elasticity";  
Science 270:293-296(1995).  
EMBL; X90569; CAA62189.1; --  
PIR; I38346; I38346.  
GO; GO:0030017; C:sarcomere; TAS.  
GO; GO:0008307; P:structural constituent of muscle; TAS.  
GO; GO:0006941; P:striated muscle contraction; TAS.  
InterPro; IPR007110; IG-like.  
DR InterPro; IPR003598; IG\_c2.  
DR InterPro; IPR004168; PPAK\_motif.  
DR Pfam; PF00047; ig; 59.  
DR Pfam; PF02818; PPAK; 53.  
SMART; SM00408; IGC2; 43.  
DR PROSITE; PS50835; IG\_LIKE; 58.  
FT NON\_TER 1  
FT NON\_TER 7962 7962  
SQ SEQUENCE 7962 AA; 883018 MW; B85240533CBAD58 CRC64;

Query Match 8.0%; Score 3313; DB 4; Length 7962;  
Best Local Similarity 21.5%; Pred. No. 18-167;  
Matches 1710; Conservative 1105; Mismatches 3287; Indels 1868; Gaps 306;

QY 10 PRFLTRPKAFVSVGKDATLSQIVGNPTPQVSWKXQOPVTAGARFRLAQDGLYRLTI 69  
DB 10 PALIPLQDVTISEGQPARFQCRVSGTDL-KVSWYKOKKIKRPSFRFTQEDYQLEI 68  
QY 70 LDALGDSQYVCRAINALGEFAAAGVLOVDRAEACAEQ-----APHLLRPT 117  
DB 69 AEAYPEDEGTYTFVANNVAGVQSSTANLSLEAPESILHERIBQEIEMEMKAPVIRKIE 128  
QY 118 SIRVREGSATRCRVGGSPRAVSNKDGRLGPDGPRVVEBELGEASALRIARAAPR 177  
DB 129 PIEVALCHLAKTCEIQSAPNRFQFKAGRIYESDKSIRSKY--ISSLEILLTQV 186  
QY 178 DGGTEVRAENPLGAASAAAALVPSDAAADTSRPGTSTAALLAHLQRREARAGAPA 237  
DB 187 DGEYTCASNEVGSVCTATLV-----TEA 213  
QY 238 SPSTGTRCTVTE--GKARLSCTVYTGPKETVMKDGQLVTEGRRHVYEDAEQNFV 295  
DB 214 YPTFLSRPKSLTTFVGGAAKTCITVTGTPVETIWMQDGAALSPSPNWI--SDAENKHI 272  
QY 296 LKILFCQSDRGLYTCTASNLVG-QYSSVLVVRPAPVFPKRLQDLVEKESATFLC 354  
DB 273 LELSNLTIDRGVYSCASNKFGADICQAEIILIDKP--HFKELEPVQSAINKVHLEC 330  
QY 355 EVDQP-STEAAMFKEETRLWASAKYIEEGTERLTVRNVSADDAVYICETPE---GS 410  
DB 331 QVEDRKTVTWSKDGQKLPFGKDYKICFEDKIATLEIPLAKLKDSTGYVCTASNEAGSS 390  
QY 411 RTVAELAVQ--GNLLRKLPRKTAVRVGDTA-MFCVELAVPGVPVHMLRNOBEWAGGRA 467  
DB 391 SCATVTVREPPSFVKVDPVSYMLPGEARLHCKLKGSPVQVTFWKNKELSESNTVR 450  
QY 468 ISAEGRHTLTISQCCLEDVGQVAFWA----GDCQTSRFRCSAPRKPLQPPVDPVKA 523  
DB 451 MYFVNSEAILDIDVKEVDSGYSCEAVNDVSDSCSTEIVIKEP--PSFKITLEPADIV 508  
QY 524 RMESVILSWSPPHGERPVTIDGVLVEKKLGTYYTWIRCHEAEMVATLTVADVABEG 583  
DB 509 R-----GTNALLQC-----EVSQGTG 523  
QY 584 NFQFRVSALNSFQSPYLEPPGTGTHLAPKLAVRTPLKAVQAVEGGEVTFSDLTVASAGE 643  
DB 524 PFEI-----S 528  
QY 644 WFLDGAALKASSVYEHCDTRHTLTIREVPASLHQAQLKFVANGTIESIRVEVRAAPGL 703  
DB 529 WFKDKQIRSSKKYRLFSQKSLVCLCIEIFSNFNSADVGEYECVWANEVKGCGCWATH----- 583

QY 704 TANKPPAAAAAR--EVLARLHEEAQLLAEL--SDQAAAATWLKDGRTLSPGPKYEVQASAGR 760  
DB 584 LLKEPPTFFVKVDDLTALGCGTQVTLQAAVRGSGSPISVTWKGQEVIREDKIKIMSSNGV 643  
QY 761 RVLVLRDVARDDAGLXECVSR---GGRIAY-QLSVQGLARFLHKDMAGSCVDVAGGPAQ 816  
DB 644 AVLIIIPDVQISFGKTYTCLAENEAGSQTSGVGLIVKEPAKIER--AELIQVTAGDPAT 700  
QY 817 FEETS-EAHVHVHMYKDGMEHLSGERFLOEDVGTGTHRLVAATVTRQDEGTY-----SCR 871  
DB 701 LEYTVAGTPELKPWKYKDGRLPLVAS-KKYRISPKNNVAQKIFYSAELHDSGQTFEISNE 759  
QY 872 VGEDSVDFRLRVSEPKVV--FAKEQLARRKLOAEAGASATLSCEVA-QAOTEVTVYKDGK 928  
DB 760 VGSSSCETTFTVLDRDIAPFTK--PLRNVDVWNGTCRLDCKIAGSLPMRVSWPKDGK 816  
QY 929 KLSSSKVCMEATGCTRLRVVQOAGADAGEYCEA-----GGORLSPHLVDKPKVPV-- 982  
DB 817 EIAASDRYRIAFVEGTASLEIIRVDMNDAGNFTCRATNSVGSKDSGALLVQEPFSPVTK 876  
QY 983 --AKDQVAHSEVQAEAGANATLSCEVAQAQAE---VMWYKDGKLSLSSLKVHAEAGCR 1036  
DB 877 PGKDVLPFGSAV-----CLKSTFGSTPLTIRWFKGNKELVSGGSCVITKEALE 925  
QY 1037 RRLVVOAGKTDAGDYSC-----ARGQVSRFLHITTEPKMFAKEQSVHNEVQAEAGASA 1092  
DB 926 SSLELYLVKTSDSGTCTCKSVNAGVVECSANLFVKEPATFVEK---LEPSQLLKGDAT 982  
QY 1093 MLCSEV-AQACTEVTVYKDGKLSLSSKVGMEVKGCTRLRLVLPQAGKADAGEYSCAAGG 1151  
DB 983 QLAQKVTGTPPIKITWPFANDREIKESKHRMSFVESTAVLRLLTDVGLTDSGEYCEAONE 1042  
QY 1152 RVSEPH---LHITPEKGVFAKEQSVHNEVQAEAGTAMLSCEVA-QPOTEVTVYKDGKLS 1207  
DB 1043 AGSDHCSSIVIVKESPYFTEK---FKPIELVKEYDVMLLAEVAGTGPFEITWFKDNTILR 1099  
QY 1208 SSKVRNEVKGCTRLRVVQOAGKADAGEYSC---BAGQVSRFQLHITTEPKAVFAEQ 1263  
DB 1100 SGRKYKTIQDHLVSLQILKFAADAGEYOQVTVNEVSGSICSAVTVLTPRPSPFIKK--- 1156  
QY 1264 VNEVTRTEAGASATLSCEVAQAQTEVTVYKDGKLSLSSSKVRIEAAQCMQLVVOQAQA 1323  
DB 1157 IESTSSLRGGTAFOATLKGSLPITVTLWTKDSDEITDDNIRMTFENNVAASLVLSGIEVK 1216  
QY 1324 DAGEYTC-----EAGQRLSPHLVDSEPKAVFAKEQLAHRKVQAEAGATLSCEVAQAQT 1379  
DB 1217 HDGRVVOQAQNDAGIQRCSSALLSVKEP-ATITBEAVS---IDVTQGDPAQLQVKSFTK- 1271  
QY 1380 EVT--WYKDGKLSLSSSKVRMEAVGCTRLRVVQOAGADAGEY-----SCBAGG 1426  
DB 1272 EITAKWFKDQGLTLGSKYKISVTDVTSILKIISTEKDSEYTFEVQNDVGRSSCKA-- 1329  
QY 1427 QRLSFLDLVABPKVYFAKEQVHREVOQAAGASTLSCEVAQAQ-TEVMWYKDGKLSLSPS 1485  
DB 1330 -RINV-LDLIIPP-SFTKKL---KMDSIKGSPFDLECIVAGSHPIQWFKDDQEISAS 1383  
QY 1486 SKVRMEAVGCTRLRVVQOAGADAGEYSC---EAGQRLSPHLHVAEPKAVFAKEQOPAS 1541  
DB 1384 EKXFSFHDNTAFLEISOLEGSTDGTCTCATNAGHNQCSGHLTVKEPP--YFVEKPS 1441  
QY 1542 REVQAEAGTSATLSCEV-AQAQTEVTVYKDGKLSLSSSKVRMEAVGCTRLRVQAEQAQAD 1600  
DB 1442 QDV--NENRVLKALVGGTAPMTIKWFKDKNELHSGAARSVMKDDTSTSLFLFAKATD 1499  
QY 1601 AGEYSCK-----AGQRLSPHLHVAEPKAVFAKEQPAHREVOQAAGASATLSCEV-AQAQT 1655  
DB 1500 SGTYICQLSNDVGTATSKATLFFVEKPPQFTKKPSV---LVLNQGQSTTFECQITGTPKI 1556  
QY 1656 EVTVYKDGKLSLSSSKVRMEAVGCTRLRVVQOAGADAGEYSCB---AGQRLSPHLHV 1711  
DB 1557 RVSNYLDGNEITAIQKHGISFIDGLATFQISGARVENSQTVYCEARNDAGTASCSELKX 1616

QY 1712 AELEPOISERPCREPLVYKHEHDIILATLATPSAATVWMLKOGVEIRRRKRHETASQG 1771  
 DB 1617 KEPTTFIRE---LKPVEVVKYSDVELECEVTGTPPFEVTLKNNREIRSSKKYTLTDVR 1672  
 QY 1772 DTHLTIVHGAQVLDGSAIYSCRYGASQDFFQVVEVAAK---FCRLLEPVCVGGELGGTVT 1827  
 DB 1673 SVFNHLITKCDSDTGEYCIVSNEGSCST-RVALKEPSPFIKKIENITTVLKSSAT 1731  
 QY 1828 LACEL--SPACAEVVRWRCNTOPRVGRQFQWVAEGP-----VRLTIVLGLRAED---AG 1876  
 DB 1732 FOSTVAGSPPIIS-ITWL-----KDDQILDEDDNVVIFSDVSVALQIRSVDNHSG 1781  
 QY 1877 EYVCESRO---DHTSAQITVSVPRVVKFMSGLSTVAEEGGEATFQCVWSPS-DVAIVW 1931  
 DB 1782 RITCQAKNESGVERCYAFLLQEP--AQIVEKAKSVDTVEKDPMTLECVVAGTPELKVW 1839  
 QY 1932 FRDGALLOPSEKFAISQSGASHSLTISDLVLEDAQITVEAE---GASS-SAAALRV--RE 1985  
 DB 1840 LKDGKQIVPSRYFSNFENNVPRIQSVMKQDSQYTFKVENDFGSSSCDAYLRVLQDN 1899  
 QY 1986 APVLFKKLEPOTVEERSVTLVLETRPWP-ELRWENATALAPQKQVVEHAGARHL 2044  
 DB 1900 IPPSTFKLLTKMDKVLGSIHMECKVSGSLPISAQWFKD-----GKEI---STSAYRL 1950  
 QY 2045 VLHNVGFADRGFFGCTPDDTKTQAKLTIVEMROVRLVRLQAVEAREQGTATMEVQLSHAD 2104  
 DB 1951 VCH-----ERSVSLVNNLEL-----EDTAN----- 1971  
 QY 2105 VDGSWTRDGLRQOQPTCHAVRGPMTLITLSGLRPEDSGLMVFAEGVHTSABLVTTEL 2164  
 DB 1972 -----YTCVSNVAGDDA-----CSGLTIVKEP 1994  
 QY 2165 PVSFGRP-LQDVV--TTEKEKVTLCBLSRPNVDVRLKDGVELPAGKMTAIAQAGCRS 2221  
 DB 1995 PSFLVKPQEQOAIPOSTVEFKALK---GTPPFKIKWFKDDVELVSGKPCFIEGTSF 2051  
 QY 2222 LTIYCEPADQGVYVCDHAQSSASVKVQGRYTLIYRRVLAEDAGBIQFVAENASRA 2281  
 DB 2052 LNLYSVDASKTQYCTCH-----VTNDVG-----SDSCTT 2080  
 QY 2282 QLRVKELPVLVRLPRLDKIAMEKHGVLCEQVSRASAOVR--WFKGSOELQPGKYELVS 2339  
 DB 2081 MLLVTEPKFKKLEASKIVKAGDSRLCEKIA-GSPEIRVWFRNEHELPAKRYMT- 2138  
 QY 2340 DGLYAKLIISDVHAEDEDTYTCADAGDVKTSAQOFFVEEQSITIVRGLQDVTVMPEAPWFE 2399  
 DB 2139 -----FIDSVAVIQWNNLSTEDSGD-----FI 2160  
 QY 2400 CETSIPSVRPPKWLIGKTVLQAGNVGLBQEGTVHRLMLRRTCSTMTGPHVFTVGKSRSS 2459  
 DB 2161 CEAQNP-----GST-----SCST----- 2174  
 QY 2460 ARLVSDIPVVLTRPLEPKTGRELSQSVLSCDFRPAPK-AVQWYKDDTPLSPKPKMSL 2518  
 DB 2175 -KVIKEPVPFSFPPIVETLKNAE-VSELCESCTPPFEVVWYKDKQLRSSKKYKIAS 2232  
 QY 2519 EQMAELRILRLMPADAGVYRQQA-GSAHSSTEV-TVEARE-----VTVTGPLQ 2565  
 DB 2233 KNFHTSIHLNVDTSIDIGYHCKAQNEVGSCTCVTKLKEPPRPFVSKLNSLTIVVAGEPA 2292  
 QY 2566 DAEATEEG-----WASFCELSHEDSEWMSLNGMPLYNDSFHEISHKGRHTLVLS 2618  
 DB 2293 ELQASIEGAQPIFVQWLEKEVIRESENI-----RITFVENVALQFAK 2337  
 QY 2619 IQRADAG-----IVRASSLKVSTSRARLEVRVKPVVFLKALDLSABEGRTLALQCEVS-D 2672  
 DB 2338 AEPANAGRYCQIKNDGGMW-ENMATLMV-LEPAVIVEKAGPMTVTVGETCTLECKVAGT 2395  
 QY 2673 PEAHVVRKDGVLGSPDKYDFLHTAGTRGLVHVHDSPEDAGLYTC---HVGSEETRAR 2728  
 DB 2396 PELSVEVYKDKLLTSSQKHKFSFNKISSRIILSVRQDAGTTFQVQNVVHGSSCTAV 2455  
 QY 2729 VRVHDJHV--GITKELKTMVELEGESCFEVLSHESADPAMWTVGKTVGSSSRFPQT 2786

DB 2456 VDYSRAVPPSFRRUKNTGGVLGASCIILECKVAGSPISVA-WPHEKTKIVSGAKYQTT 2514  
 QY 2787 QGRKYLIVREAAAPSDAGE---VWFSVRGLTSKASLIVRERPAALIKLEDOQWAPGED 2843  
 DB 2515 FSDNVCTQLNSLSDSDMGNYTCTVAANVAGSDECRVILVQPPSPVKEPEPELVLPGN 2574  
 QY 2844 VELRCELSRAGTP---VHMLKDKARKQKQYDVCEGTNAMLVIRGASLKADGETVCEV 2900  
 DB 2575 VTFTSVI--RGTPPFKVNFRGARELVKGDRCNIYFEDITVABELEFNIDISQSGEYTCV 2632  
 QY 2901 EASKSTAS---LHVEKANCFTTELTNLQVEE-KGTAVFTCKTEHPAATVTRWKGLEL 2955  
 DB 2633 SNNAGQASCTTRLFVKEPA-AFLKRLSDHSVBPCKSIILESTYTGTLPISTVWKKDGFNI 2691  
 QY 2956 RASGRHQPSQEGTLTSLTISALEKADSDTYTCDIGQAQRAQLLVQGRVHIIE----- 3009  
 DB 2692 TTSEKCNIVTTEKTCILEILNSTKRDAGQVSCBI---ENEAGRDVCCALVSTLEPPYFT 2748  
 QY 3010 DLEDVDVQSGSSATFECRISPANYEPVHWFLDKTPHANE-----LNEIDAQPCGYHVL 3063  
 DB 2749 ELEPLEAAGVDSVSLQCVAGTPEITVSWYKGTUKLRPTPEYRTYFTNNV-----ATL 2801  
 QY 3064 TLRQALKDSTGIYFEA---GDORASAAALRVTEK--PSVFSRELTDAITIEGEDITLYC 3117  
 DB 2802 VFNKVNINDSGEYTKCAENSIGTASSKTVFIQEROLPPSFARQLKDIEQTVGLPVTLC 2861  
 QY 3118 E-TSTCDIPMCWTXKDKITLRSARCOLSHEHRAQLLITGATLQDSGRYKCEAG---GAC 3173  
 DB 2862 RUNGSAPIQVCMYRDGVLLRHENLQTSFVDNVATLKILQTLDSHSGQYSCSASNPLGTA 2921  
 QY 3174 SSSIVVHARPVR---FOEALKDLEVEGGAATLRCVLSSVAAPVK--WCYGNVNLRRP 3227  
 DB 2922 SSS-ALUTAREPKSPFFDIKPVSDIVIAGESADFECHVTG-AQPMRITWSKDNKEIRP 2979  
 QY 3228 DKYSLRQEGAMELVVRLNRPODSGRYSC---SFGDQTSATLIVTALPAQIFOKLR-N 3282  
 DB 2980 GNYTITCVNGTPHLRILKVGKDSGQYTCQATNDVGMKCSAQLSVKE-PPRFVFKLEAS 3038  
 QY 3283 KEATEGATATLRCELSKT--APVEMKSGSETLRDGRYCLRODGMCELOIRGLAMVDA 3340  
 DB 3039 KVAKGESIQLECKISGSPFIKSHFRNDSEHESKYNMFSINSVALLTINEASAEBSG 3098  
 QY 3341 EYSCVC---GEERTSASLIRPMPAHFTIGRLRHOESIEGATATLRCELSKAAPVE--WR 3394  
 DB 3099 DYICEAHNGVDASCSALTIVKAPPV-FTQKPSVGLKAGSDVILQCSIGSTPPEVVMV 3157  
 QY 3395 KGRESLRDGRHSLRQDGAVALCELQICGLAVADAGEYSCVC---GEERTSATLTVKALPA 3450  
 DB 3158 KDKQVNRNKKPKITSKHFDTNLHNLNLEASDVGEYHCKATNEVGSDDTCSGVKFE-PP 3216  
 QY 3451 KTEGLRNEAEVAGATAMLWCELSKVAPVE--WRKGPENLRDGRYILRQEGTR----- 3502  
 DB 3217 RFVKLSLSTLIGDAVELRAIVEGFPISVVW-----LXDRGEVIRESENTRISFDN 3270  
 QY 3503 -CELOQCGLAMADAGEYLCV---CGERTSATLIRALPARFIEDVKNQOAREGATAVL 3557  
 DB 3271 IATLOIGSPSEANSKGIYCOIKNDAGMRECSAVLTVLE-PARIIEKPEPMTVTTGNPFAL 3329  
 QY 3558 QCELSAAV--EWRKGSSETLRDGRYSRLRQDQTKCELQIRGLAMADTGEYS----- 3607  
 DB 3330 ECWVTGTPELSAKFKDGHLSADSXGHTTIFINKVASLKI PCAEMSDKGLYSFEVKNVSG 3389  
 QY 3608 ---CVCGQERTSAMLTVRALPIKFTGLERNEBATEGATAVLCELSKAPVE--HWKHE 3662  
 DB 3390 KSNCT-----VSVHSDRIVPSPFIRKLVDNAILGASVLECRVSGSAPISVGFODGN 3444  
 QY 3663 TLRDGRHSLRQDGAVALCELQIRGLVAEDAGEYLCV---MCKERTSAMLTVRAMPSKIE 3718  
 DB 3445 EIVSGPKCSSSENVCTNLNLSLSPDTGTYTCAANVAGSDECSAVLTVQEPSP-FEQ 3503  
 QY 3719 GURNEEATEGDTATLWCELSKAAP--VEMRKGHETLRDGRHSLRQDGRSCELQIRGLAV 3776

3504 TPDSEVFLPGMSLTFTSVIRGTPPFKVKWFGSGRELVPGESCNISLEDFVTELELFEVQP 3563  
3777 VDAGEYSCVCGQERTSALTIVRAL---PARFIEDVKNQAREGATAVLOCELSKAAP--V 3831  
3564 LESGDYSLVNTDAGSASCETHLFVKEPATFKRLADSFVETGSPVILEATYTGPPISV 3623  
3832 ENRKGETLGRGDRYSLRGDRFCRLOHGLSVADTGEYCV---CGOERTSATLTVRA 3887  
3624 SHIKZEYLLISQERCSINTWEXSTILELESTIEDYAYQYSLIENEACQDICEALSVLE 3683  
3888 POPVREPIQSOAEAGSTATLOCEL-SEPTATVVMVSGGLOQLQANGRRRPLQOCTAEL 3946  
3684 P-PYFTEPLEHVEAIVGEPATLOCKVDGTPFEIRISWYKEHTKRLAPAYKMQFKNVASL 3742  
3947 VLQDLQREBTDGVTG---TCGSOATSATLTWTA--APVRELELHQHVEDEGGTAHLCC 4000  
3743 VINKVDHSDVGEYSCADNSGAVASSAVLVTKAKKLPPFPARKUDVHETLGFPPVAFEC 3802  
4001 ELSAR--GASVEVRKGSLOLFPCKAYQMVQDGAALVVRGVEQEDAGDYTCDTGHTQMA 4059  
3803 RINGSEPLQVSWYKDGVLKDDANLQTSFVHNVALTQILQTDQSHIGVNCASNPLGTA 3862  
4060 SLVSR-----VPRPKFTRLOSLEOETGDIARLCCQLSDAESGANVQWLKEGVELHAG 4112  
3863 SSSAKLILSEHEVP-PFFDLKPVSDVLALESGETFKCHVT-GTAPIKITWAKDNREIRPG 3920  
4113 PKYEMRSQATRELLIHOLEAKDGTGEYAC---VTGGQKTAASLRVTPBEVTIVRGLVDA 4168  
3921 GNYKMTLVENTATLTVLVYKVGKDGAGQYTCYASNIAGKSCSAQLGVQBPFFIKLEPSR 3980  
4169 EYTADEBVEFSCEVSRAGATGVQWCLQGLPLQSGNEVTEVAVRDGRHILRLKGVTPEDAG 4228  
3981 IVKQDEFTREYCKIGGSPKIVLMYKDETEIQESSKFRMSFVDS--VAVLEMNLVSDESG 4039  
4229 --TVSPH--LGNHASSAQLTVRAPEVTILEPLQDVQLSGQDASQCRLSRAGSQEAWA 4284  
4040 DYTCEAHNAGASASSTSLKVEPPFRKPP-HPIETLKGADVHLECELOQTFFPHVSNY 4098  
4285 LGGVPLOANEMNDITVEQ--GTLHLTLHKVTL-----EDAGTVSPHVGTCSSEAL 4334  
4099 KDKRELRSKKYKIMSENFELTSHILNDAADIGEYCKATNDVGS-----DTCVGSIAL 4153  
4335 KVTAKNTVYRGLNVALEGGELFECQLSQEVAHAHTWLLD--DRPVRTSENAEVVFEEN 4393  
4154 K--APRFRVKJUSDITVTVGKEVQLOTTIEGAEPIVSVWFKOGEIVREDNIMWISYEN 4211  
4394 GLRHLLLLKMLRPQD---SCRVTFLAGDMVTSAFILTVRGWRLILEPL-----KNAAV 4443  
4212 --IATLQFSRVEPANAGKYTCQIKNDAG--MQSCFAT----LSVLEPATIVEKPESIKV 4262  
4444 RAGAQAARFTCTLSEAVPVGBASWYINGAAVQPDSDTWVTADGSHQALLRSAPHAGE 4503  
4263 TTGDTCTLECTVA-GTPELSTKWFQKGL-TSDNKYKISFFNKVSGSLKIINVAPSDSGV 4320  
4504 VTFACRAV---ASARLTVLGLPDPPEDA-----EVVAHSHVTTLWS 4543  
4321 YSEVQNPVUKDCTASLOVSDRTPPSFTRKLKETINGLSGSSVMVECKVYGSPPISVSM 4380  
4544 AAPMSDGGGLCGYRVEVEKATGQWRLCHELVPGPECVVDGLAPGETYRFRVAAPVPGV 4603  
4381 ---FHEGN-----EISSGRKYQTLTDNTC--ALATVNMLESDSDGYTCIATNMAG 4426  
4604 AGEPVHLPOTVRLAEPPKPYPPQSPAPESRQVAAGEDVSILELVVAEAG-EVWHKGMER 4662  
4427 SDE--CSAPLTVR--EPPSFV---QRPDPMDVLTGNTVFTTSIVKGTDPFVSFWFKGSSE 4479  
4663 IQPGGREPVVSQGRQWMLVKGTAEQDQYHCGLAQGSICPAAATFOVALSPASVDEAP 4722  
4480 LVFGDRCNVSLDSVABLELFDVDTSGSGEYTCIVS-----NEAG 4519  
4723 QPSLPPPEAAQEGDLHLLWEALAR--KRRMSREPTLDSISELPEDGRSQRLPQEAEEVAP 4780  
4520 KASC-----TTHLYIKAPAKFVRK-----LNDYSIEKCK-----P 4549

4781 DLSEGYSTADELARTGDADLSHTSSDDESAGTPESLVTVLKAGRPOTSPASKVGPAPAA 4840  
4550 LILEGTF-----GTFPISVTKONG----- 4570  
4841 PSVKPQOQOPEPLAARPPPLGDLSTKOLGDPMDKAAVKIOAAGKGVKVRKEMKQOQGPME 4900  
4571 INVTSPQRNITTEKSPILEPSTVEDAGQNCVIE-NASGKD-SCSAQILILEPPYPF 4628  
4901 SHTFGDTAEAOVGDALELCVAVASKADVARAWLKDGVEL---TDGRHHHIIQLQDGTCSLL 4957  
4629 VKQLEPVKVSVDGASALQCOLAGTPEIGVSWYKGTDLRFTTYTKM-----FRNVATLV 4684  
4958 IAGLDRADAGCYTCQVSNKFGQVTHSACVVVSGSESEAESESSGGELEDDAFRRARLHRL 5017  
4685 ENQVDINDSGEYICKAENSVEGVSASTFLTVOEQ-----KLPPSESRQLRDVQE- 4733  
5018 PRKSPAEVDEBELFLSADGEPRAEPEDAPWQYREDEHEFICIRFEALTARQAVTEFQE 5077  
4734 -----TVGLPVVFDCAISG-----SEPISVSWYKDG-----KPLKDSNNVQTSFLD 4774  
5078 MFATLGI-GVEIKLVEGPRRVMCISKETPAVPVPEPLPSLLTSDAAPVFLTELQNE 5136  
4775 NTATLNFKTRSLAGO-----YSCTA-----TNPIGS-ASSARLILTEGNPP 4818  
5137 VQD-----GYPVSDCVVTGQMPMSVRWFKDGKLLBEDDHYMINEDQOQGHQJLI 5186  
4819 FFDIRLAPDAVVGESADFECHVTGTQPIKVSWAKDSREIRSGKYQISYLENSAH-LTV 4877  
5187 TAVVPADMGVYVCLAENSMGVSTKAEALRVLDLTSTDYTAADATESSESYFSAQYLSRE 5246  
4878 LKVDKSGQYTCVAVNEVKDSCATQINI----- 4909  
5247 QEGTESTTSGQLQPVVEYLRDLQVAPGTRELAKFQL--KVGYPAPALYKFGQOQPLTAS 5304  
4910 RLIPPSFTK--RLSETVEE-----TEGNSFKLEGRVAGSQPIVAVYKNNIEQPT 4958  
5305 AHIENTGKILHTLEIISVTRDSQYAAVTSNAGAAVYASARLLVGPDEPEEKPSADV 5364  
4959 SNCEITFKNTVLVQVAKAGNDAGLYTCKVNDAGSALCTSSIVIK---EPKPPVFDQ 5015  
5365 HEQLVPPRMLERFTPKVKKGSSITFSVKVEGVPVTVHVLREBAERGVLWIGP-DTPGY 5423  
5016 H-----LTFVTVSEGYVQLSCHVQSEPIRIQMLKAGRE-----IKPSDRCSF 5059  
5424 TVASAAQSHSLVLLDVGRHQGTTCIASNAAGALCASLHVSGLPKVBEQEKVKEALI 5483  
5060 SFASGTA--VLELRDVAKADSGDYVCKASNVAGSDTTKSVTKDKPAPAPATKKAADV 5117  
5484 STFLOQTTOATSAQGLETAGF-ADLGGORKEEPLA-AKEALGHLSLAEVGTETEFLOKLT 5541  
5118 RLFFVSPQSIRVVEKTTATFIKAVGG---DPIPNVYKWKGRQLNQGRVPIHQ--- 5170  
5542 QITEMVSAKITQAKLVPGGSDSDSKTPS-----ASPRHGRSRPSSSIQ---ESSSES 5592  
5171 -----KGDEAKLEI-----RDTTKTDSGLYRCVAFNEBHEGESNVNQLQVDERKKQEK 5217  
5593 EDGDARGEIPDIYVVTADYPLGAEQDAITLREGQYVEVLDAAHPLRWLVRTKPTKSPS 5652  
5218 IEGDLRA-----MLUKKTPILKKGAGE-----EEDIMELLKNVDP----- 5253  
5653 RQGWVSPAYLDRRLKLSPEWGAEEAPFPGAEVSEDEYKARLSVSIQELLSQOAF--VE 5710  
5254 -----KEYEKYARMYG-ITDFRGLLQAFELLK 5279  
5711 ELQFLOSHLOLHERCPHVPIAVAGQKAVIFRNVVDIGRHSFLQELQOCTDDDDVAMC 5770  
5280 QSQEEETHRLS-----IEEIERSERO----- 5300  
5771 PIKQAAPEQYLEFLVGRVQAEVWVSTAIQFYKYAAEALLAGDSPPPPPLOHYLE 5830  
5301 -----EKEFEELVSTIQEL-----SQTEP----- 5320

QY 5831 QPVERVQRVQALLKELIRNARNRQNCALLBOAYAVVVSALPORAENKLVHLSMENYPGTL 5890  
Db 5321 -----VTLIKQ-IEQNTVLKONDAVE-----IDIKINYP----- 5349  
QY 5891 EALGEPHQHFTWEGAGPMPKGNHRHVFLEFRNLHVICPDRDRTDTVSVFRNM 5950  
Db 5350 ----- 5350  
QY 5951 MKLSSIDLNDQVEGDDRAPEVQWEREDSVKYLQARTAIKSSW-VKEICQIORLALP 6009  
Db 5351 IKLSWKYKTELEPSDK-PEISIDGD-----RHLRVQNCQLKQDQGNVRLVCGPHIASAKL 5405  
QY 6010 VWRPDPFEBELADCTAELGETVKLACRVT--GTPKPVISWYKCKAVQVDPHLLIEDP 6066  
Db 5406 TWIEFAWEHLQDVTIKESQT-----CTMTVQSVNVKSEVFRNGRIILKQGRH-KTEVE 5460  
QY 6067 DGSCALILDSLTGVDGQYMCFAAAGNCSTLGLKLVQVPP-REFNVKVRASPFVEGEDA 6125  
Db 5461 HKVHKLTITADVRAEDQGYTC-----KYEDLETSAELEAEPIQFTKRIQNVSEHQA 5516  
QY 6126 QFTCTIEGAPYQIWKYKDGALLTGNKFOTLSEPRSGLLVLVIRAAASKEDLGLYECELV 6185  
Db 5517 TECEV-SDDAIVTWYKGTLETESQYFNFRDGRCH--YMTIHNVTPDDEGVY--SVI 5571  
QY 6186 NRL---GSARASAEIRIOPMLQAOEQCHREQLVAAVEDTTLERADQVTSVLKELLGPK 6242  
Db 5572 ARLEPRGEARSTAEVYL-----TTKEIKLELKP- 5599  
QY 6243 ACPSTGDLTGCP-CRCPALQETGQPPVTGTSEAPVPRPVQPLLHGEPEQPEA 6301  
Db 5600 ---PDPSRVPIPTMPTRAVPEEI---PPVV-----APPVLLLETP-----BEKPP- 5643  
QY 6302 IARAQEWTPVIRMEGAAMPAGCTGELLMDVHSHVVRETTQRTYTYQAITDHTAAPPMSOV 6361  
Db 5644 -----PKIE-----VTKAVKDAKV-----VAKKEM-- 5668  
QY 6362 TIEDVOAQGTGAQAEIIEGDPQBSVTWKDSVOLVDSTLSQOQEGTTVSLVLR--HV 6419  
Db 5669 -----TPREEIVKPPPTTILPAKPEIID---VSSKAEEVKIMTIRKEV 5713  
QY 6420 ASKAGVYTCIAQNTGGVCLCAELLVLGGDNEP--DSEKOSHRRKLHS-FYEVEKEIGR 6476  
Db 5714 QKEAVY-----EKQAVHKEKVFIETSEEPYDELEVPTPEFPQYEPDDEYE 5767  
QY 6477 GVGFVKRVQKGNKILCAAKFIPILRSRTAQAYREDILAAHSLPLVTGLDQFETRT 6536  
Db 5768 EIKVEAKKEVHEWE-----EDPEEGQYVERE-----EGYD--- 5799  
QY 6537 LILILELCSBELDLRYKGVVTEAEVKNVIOQLVEGLHLSH-----GVHLHLIKP 6590  
Db 5800 -----EGEEWEBA\*CEERIVQKEVYEE-----SHERKVPKVPKAPP 5841  
QY 6591 SNILMVHPAREDKICDFQAFQNTIPAELOFSQYSGSPFVSPEITIQQNPVSEASDIWAG 6650  
Db 5842 PPKVIKPVIEKIE---KTSRMEEEKVQTKV--PE-VSKIVPQKP----- 5883  
QY 6651 VTSLSLTCSPFAGESDRALLNVLEGRVSWSPMAAHLSEDAKDFIKATLQAPARP 6710  
Db 5884 -----SRTVQSE-----VIEKVPAVHTKGMVISEEOMFFASHTEEVSVTVP 5927  
QY 6711 SAAQCLSHPWFLKSPABEAHFINTKQLKLLARSWQSLMSYKSLVMSRIPELLRGP 6770  
Db 5928 EV-----QKEIVTEKHV--AVSKRVEPP 5950  
QY 6771 PDPSLGVARHLCRTGTGSSSSSSSSSNDLEAPFAKSL-PPSPVT----- 6815  
Db 5951 PKVPEL-----PEKPAPEVAPVPIPKVVEPAPKVPVPPKVPPEEKPP 5995  
QY 6816 -----HSPLLHPRGLRPSASLPEEAESERSTAPAPPASPEGAGPPAAQGGCVPRHSVI 6870  
Db 5996 VPVPKKEAAPPKVPPEKVPPEKIPVPAKKEAPPKVP-----PEVQKGVVTEKIE- 6050  
QY 6871 RSLFVHQAGESPEHGA-----LAPGSRHRHARRHLLKGYIAGALPGLREPLMEHRV- 6923

Db 6051 --TIVTQRESPPPAVPEIPKKKVPPEKVPPEKVEEVPVPPKVPALP--KKVPEEKVA 6106  
QY 6924 -----LEEEAAREEQATLLAKAPSFETALRLPASGTHLAPGH--SHSLH 6966  
Db 6107 VPVPVAKKAPPRAEYVSKTVVEEKFAVEEKLSPAVQEVETREVEAEBSWSEEE 6166  
QY 6967 DSSP-----TPRPSSACGE-----AQLPSPSGGAPIRDMG 6999  
Db 6167 EGVSIYVREEREEREAEAEVTEYVWEPEEYVVEEKLHIISKRYEAPPA-----EVT 6220  
QY 7000 HPGSK-----QLPSTGGHPTAQ--PERSPDPMGQ--PAPFCHPKQGSAPQEGCSHPA 7052  
Db 6221 ERQEKILVLPKIPAKIEPPPPAKVPEAKKIVPEKKVPAPV--PKKVPVPPKVPPEEK 6278  
QY 7053 VAPCPGSPFPSPGCK-EAPL-----VPSPPFLGQPOAPPAPAKASPLDSKMG- 7099  
Db 6279 -KVPPEKVPKVIKMEELPAKVTBKMQITQEEKVLVAVTKKEAPPXARVPEEPKRAV 6337  
QY 7100 PGDISLPGRPKPCPCSPGASOASSQYSSLVGSSQVGTBPGSLDAEGHTQE--AED 7157  
Db 6338 PEEKVLKLPKREE-BPPAKVTEFRKRVVKEEKVSTEAPKBPQPIKEVTIMEEKERAYT 6396  
QY 7158 LSDSTTLPQPOQATWRKFSLGGRGYAGVAGYGTFAFGDAGGMLGQGPMMARIAMAV 7217  
Db 6397 LEEEAHSVQREEREYEEYDYKEFEYEPTBEYDQY----- 6433  
QY 7218 SQSEEBEQSEARABSOEQOQARABSLPQVSARVPVPEGRAP-----TRSSPE--PT 7269  
Db 6434 --EYEREYEREEHEEYITEPKIPVKVPPEEVPPTKPAKPAKVLKAVPEEKVPV 6491  
QY 7270 PWEDIQVSLVQIRDLSDGDAEAADTISLDSVDPAVNLSDLYDKILPPEFMIFRKVP 7329  
Db 6492 P----- 6494  
QY 7330 KSAQPPSPMAFEELAEFPPTWMPGELGPHAGL-----EITESEEDV--DALLA 7379  
Db 6495 KKLKPPPK-VPEEPKVPFEKI-----HISITKREKEQVTEPAKVPKPKRVVA 6544  
QY 7380 EAAVGRKRKWSPPSRSLPHFPGK--HLPLD-BPAELGLRERVKASVEHISRL----- 7429  
Db 6545 EEKVPVPRKEVAP-----PVRVPEVPALEPPEVAFEVEV---VTHVEEVLVEEVEEY 6594  
QY 7430 -----KGRPEGLEKEGPPPKKP 7446  
Db 6595 IHHEEFITEEVEVVPVVKVPEVPRKVPPEKVPVPPKKEAPPKVP 6644

## RESULT 12

Q8NHNO Q8NHNO PRELIMINARY; PRT; 1020 AA.  
AC Q8NHNO;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Obscurin (Fragment).  
GN OBSCN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Young P., Ehler E., Gautel M.;  
RT "Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere assembly";  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
DR EMBL; AJ314908; CAC85755.1;  
DR GO; GO:0005524; F-ATP binding; IEA.  
DR GO; GO:0004812; F:RNA ligase activity; IEA.  
DR GO; GO:0006418; P:amino acid activation; IEA.  
DR InterPro; IPR003599; Ig.

Qy 6287 -POPLHGEPE-----QEPAIARAQEWTPVIRMEGAAM--PGAGTG 6325  
Db 720 GPQPKWVEVEETIEVRVKMGPOGVSPTE--VPRSSGHFLPLPGATPG 767

RESULT 13

Q7Z120 PRELIMINARY; PRT; 8081 AA.

AC Q7Z120;  
DT 01-OCT-2003 (T-EMBLrel. 25, Created)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
DE Uncoordinated protein 89, isoform b.  
UN 89.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Wilson R.;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Du Z., Le T.T., Wilson R.;  
RT "The sequence of C. elegans cosmid C09D1.";  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Wilson R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF003131; AAP68958.1;  
SQ SEQUENCE 8081 AA; 894245 MW; 67C804953CF62228 CRC64;

Query Match 7.7%; Score 3182.5; DB 5; Length 8081;  
Best Local Similarity 21.1%; Pred. No. 1.1e-160;  
Matches 1891; Conservative 1190; Mismatches 3350; Indels 2549; Gaps 375;

Qy 27 ATLSQIVGNFTPOVSWKQDQPVVAGARFRLAQDGD-LYRLTILDALGDSQYVCRAR 85  
Db 565 AKMCKLVGTGYPDITWYKDDVQLHEDERTFYSDDEGDFAMTIDPVQVTDTCRYTCMAT 624

Qy 86 NAIGEAFNAV---GLQYDAEACAEQAPHELLRPTSTRVREGSEATFRCVGGSPRAVS 142  
Db 625 NEYGOASTSAFFRVLKVEKEA---PPAFVTKLRDECKEGDVIDECEVEGWPEPELV 680

Qy 143 WSKDGRILGEPDGPVRVVEELGEASALRIRAAAPRDGTYEVRAENPLGAASAAAALVVD 202  
Db 681 WLVDQDPLRPSHDPLQYD--GQPAKLEIRDAQDDTGVTYVKIQNEFGSTESKAELFVQ 738

Qy 203 SDAADTASRPTSTAAALLAHQRREARAEAPSPSTGTRCTCTVTEGKARLSCVIT 262  
Db 739 ADPKNHVAP-----EFQATIEYVECD-----EGEEVRFKSVIT 772

Qy 263 GERKPEVTKKDGQLVTEGRR-HVVYEDAOENFVLKILFCQSDR---GLYTCTASNLVG 318  
Db 773 GDNPEIIFWINGKLPSESEKVKFISBDG-----ICLTIKDVTFRHFGVMVTCQGSNRLG 827

Qy 319 QYSSVLVVRPAVP--FKRLQDLVREKESATFLCEV---POPSTEAAWFKBEETRLW 373  
Db 828 SASCDGRLKVRVPPAPPTFNKPLEDKTVQEKSTVVFYDVSGWPEPTLTFTLCKELK-- 885

DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003598; IG c2.  
DR InterPro; IPR001849; PH c2.  
DR InterPro; IPR000219; RHOGEF.  
DR InterPro; IPR001452; SH3.  
DR InterPro; IPR001412; tRNA-synt\_1.  
DR Pfam; PF00047; IG; 2.  
DR Pfam; PF00169; PH; 1.  
DR Pfam; PF00621; RHOGEF; 1.  
DR SMART; SM00409; IG; 2.  
DR SMART; SM00408; IGC2; 2.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00325; RHOGEF; 1.  
DR PROSITE; PS00178; AA TRNA\_LIGASE\_I; 1.  
DR PROSITE; PS0010; DH 2; 1.  
DR PROSITE; PS0835; IG LIKE; 2.  
DR PROSITE; PS0003; PH DOMAIN; 1.  
KW Immunoglobulin domain.  
FT NON\_TER 1  
SQ SEQUENCE 1020 AA; 112239 MW; ALFABD68B3F1948B CRC64;

Query Match 7.9%; Score 3270; DB 4; Length 1020;  
Best Local Similarity 84.9%; Pred. No. 6.4e-167;  
Matches 654; Conservative 14; Mismatches 54; Indels 48; Gaps 9;

Qy 5601 IFDIYVVTADYLPGLAEQDAITLREGQYVEVLDAAHPLRLVTRTKTSSPSRQGWSPA 5660  
Db 1 IFDIYVVTADYLPGLAEQDAITLREGQYVEVLDAAHPLRLVTRTKTSSPSRQGWSPA 60

Qy 5661 YLDRLRLKSPENGAEAPFPEGEAVSEDEYKARLSSVTOELSSQAFVEELQFLQSHHL 5720  
Db 61 YLDRLRLKSPENGAEAPFPEGEAVSEDEYKARLSSVTOELSSQAFVEELQFLQSHHL 120

Qy 5721 QHLRCPHYPIAVAGQKAVIFNRVDRIGRHFSSFTQEQDCTDDVAMCFIKNOAFAEQ 5780  
Db 121 QHLRCPHYPIAVAGQKAVIFNRVDRIGRHFSSFTQEQDCTDDVAMCFIKNOAFAEQ 180

Qy 5781 YLEFLVGRVQASVVTSTAIQSFYKYABEALLAGDPSPPPPLQHYLEQVVERVQYQ 5840  
Db 181 YLEFLVGRVQASVVTSTAIQSFYKYABEALLAGDPSPPPPLQHYLEQVVERVQYQ 240

Qy 5841 ALLKELIRKARNRQNCALLEQAYAVSALPORAENKLVSLMENYPTGLAIGPIRQG 5900  
Db 241 ALLKELIRKARNRQNCALLEQAYAVSALPORAENKLVSLMENYPTGLAIGPIRQG 300

Qy 5901 HFTVWEGAPGAPWGHNRHVPFLFRNLVICKPRDSRTDTSVYFRNMKLSLIDLND 5960  
Db 301 HFTVWEGAPGAPWGHNRHVPFLFRNLVICKPRDSRTDTSVYFRNMKLSLIDLND 360

Qy 5961 QVEGDDRAPEVWQEREDSVRKYLLOARTAIKSSWYKEICGQQRLALPVRPDPFEEL 6020  
Db 361 QVEGDDRAPEVWQEREDSVRKYLLOARTAIKSSWYKEICGQQRLALPVRPDPFEEL 420

Qy 6021 ADCTAELGETVKLACRVGTGPKVVISWYKDGKAVQVDPHILIEDPDGSCALILSLTGV 6080  
Db 421 ADCTAELGETVKLACRVGTGPKVVISWYKDGKAVQVDPHILIEDPDGSCALILSLTGV 480

Qy 6081 DSGQYWCFAASAAGNCSTLGIQVPPFRFVNKVRASPFVEGEDAQFTCTIEGAPYQIR 6140  
Db 481 DSGQYWCFAASAAGNCSTLGIQVPPFRFVNKVRASPFVEGEDAQFTCTIEGAPYQIR 540

Qy 6141 WYKDGALLTGNKFTLSEPRSGLLVLVIRAASKEDGLGYECELVNLRLGSASAEFLRQ 6200  
Db 541 WYKDGALLTGNKFTLSEPRSGLLVLVIRAASKEDGLGYECELVNLRLGSASAEFLRQ 600

Qy 6201 SPMLQAOEQCHRLQVAIVEDTLER-----ADQEVTSVLKRLGPKAPGSPGDLT 6252  
Db 601 SPMLQAOEQCHRLQVAIVEDTLER-----ADQEVTSVLKRLGPKAPGSPGDLT 659

Qy 6253 GPQPCPRG-----APAL-----QETGSP--PVTGTSEAPVPPRYV---- 6286  
Db 660 SPSPCSASPLRPGLLAPDLLYLPAGACPRRPEAPGKPVWVPLYVTEAEASPALPGLS 719







Db	2591	TFPVKWIYIGTQVLCATEDTEISNANGVY-TMKIHGVTADMTGKIKCVAYNKAGEVSTBEG-	2648
Qy	2372	FFVBSQSIIVRGLOQVTVMEPAPAMFE---CETSI-----PSVRPPK	2411
Db	2649	-----PLKVAPIPVEFETSLCDATCREGDTLKLRAVLLGPEPVW---S	2690
Qy	2412	WLLGKTVLQAGNVGLEBEGTVHRLMLRRTCSTMGPVHFVTV---GKSRSASRLVWSDI	2467
Db	2691	WTVNGKKLEESQIKIHSEKGYTYTIIKDITDISGVQVVCALINEYGAATSEATL-----	2745
Qy	2468	PVVLTRPLEP-----KTG-RELQSVVLSCDPRPAKPAVOWYKDDTFLSPSEKFK	2515
Db	2746	-LVLPRGPPDFLEWLSNVRAARTGTWVHKVVFTGDPKP---SLTYINNKEILNSDUYT	2801
Qy	2516	MSLEQMAELRILRLMP-----ADAGVVRCOA-----GSAHSSTEVTVARE	2557
Db	2802	IVTDDKTSLTITNSFNPVHVGEIICKAENDAGEVSCFANMITYTSDMFSSESBAQAE	2861
Qy	2558	-----VTVTGP-----LQDAEATEEGWASFSCBLSHEDDEE---VE	2589
Db	2862	FVGDDLTDESEUREEMHRTPTVPMAKPTIKDKAKKGSHSAVFCVV---PDTKGVCK	2919
Qy	2590	WSLNGMPLYNDFHISHKGRHT-----LVLKSIORADAG-----IVRASSILVK	2634
Db	2920	WLKDKKEI-----ELIARIVQTRTGPBGHITQBELVLNVTTPEDAGKYTCIVENTAGDK	2973
Qy	2635	STSAARLEV-----RVKPVVFLKALDLSAEBRGTLALQCEV-SDPEAHVVRKDGQV	2685
Db	2974	TCEATLTVIESLEKSEKAPFIVALQDKTKTSEKVVLECKVIGEPKPKVSWLHDKNT	3033
Qy	2686	LGPSDKYDLHAGTRGLVHVDSPEDAGLYTCHV-----GSEETRAVRVHDLHVGITRK	2741
Db	3034	I-TQESITVESVEGVERVITISSELSHQKYTCIAENTEGTSKTEAFLTVQGEAPVFTKE	3092
Qy	2742	LKTMVLEGESCFEVLVSHESASDPAMTVCGKT-----VGSSSRF-----QATROGRK	2791
Db	3093	LQNKELSIGEKLVLSCVSGSQPHVDFYSFSETTKVETKITSRRIALIEHDQTNHWM	3152
Qy	2792	YI-LVVRREAPSDAGEWFSVRGLTSKASLIYRERPAALIKPLEDQWVAPGEDVLRCEL	2850
Db	3153	VISQITKEDIVSKALATNSIGTATSTKITTKVEAPVEQGLKTSVKEKEIKWEMVKV	3212
Qy	2851	SRAGTPVHVKOKAIRKOKYDVV---CEGTWAMVIRGASLKADAGEVTCVEASKSTAS	2908
Db	3213	GGAPVDFWKDDKPVSEDNHEMKNPETGFTLVVQKAAITDAGKYTAKASNPAGIAE	3272
Qy	2909	LHVEERKAN-----CPTBEL---TNLQVEEKGAVFTCKTEHPAATVWRKGLLELRASGK	2960
Db	3273	SSAAEAVTQSLKPTVRELVTTEVKINETATLSVTVKGV-FDPSVEMLKQDQPVQDSS	3331
Qy	2961	HQPSQ-EGT-TURLTISALEKADSDTYTC-----DIGQASRAQLLVQGRV--HIIEBLE	3012
Db	3332	HVIKVEGSGYSITIKDARLEDSGKYACRATNPAGEAKTEANFAVVKNLVPPFVEKLS	3391
Qy	3013	DVDVQEGSGATPCRISIPANYEPVHFLDKTPLHANELNEIDAQPG-GYHVLTLRQLALK	3071
Db	3392	PLSVKEKESTLSVKVGVGTPPEVSEVNFKDDTPISIDNVHVIQKQAVGFSFSLTINDARQ	3451
Qy	3072	DSGTIYF-----BAGQORASALRVTEK---PSVFSRELTDAITEGEDTLVLCSTDCDI	3124
Db	3452	DVG-IYSCRARNEAGEALATANPGIIRDSITPPEFTOKLRPLEVREQETLTLKVTVIGTPV	3510
Qy	3125	P-MCWNKDKGTLRGSARCQLSH-----EGHBAQLLITGATLQDSGRYKC---BAGGA	3172
Db	3511	PNVWFKDDKPI-----NIDNSHIFAKDESGGHIT-ITIQARGEDGVVTCATNEAGEA	3565
Qy	3173	CSSSIVRVHAR---PVRFOBALKDLEVLGGAAATLRCLVSSVAP-VKWCYGNVNLRPD	3228
Db	3566	KTTANNAVQBEIEAPL-FVQGLKPYEVEQGPALVVRVEGKPEPEKPKFKDGPVIAIDN	3624
Qy	3229	KYSLRQBG---AMELVVRNLRPQDSGRYSC-----SFGDQTTSATL-----TVTAL	3272
Db	3625	QHVIERKKGNGSHTLVIKDTNNADFGKYTCQATNKAKGDETVGELKIPKYSFEKQTAEBV	3684
Qy	3273	PAOFIGKLRNKEATEGATATLRCELSKTA--PVEMRKGSETLRDGDYCLR---QDGAMCE	3328
Db	3685	KPIFIPLEKETFAVEGDTVVLECKVNKESHPIKFFKNDQPVIEIGHQMLVELEDGNI-K	3743
Qy	3329	LQIRGLAMWDAABYSC-----VCGEERTSASLTI-----	3357
Db	3744	LTIQNAKKEDVGAYRCEAVNVAGKANTNADLKIQFAAKVEEHVTDSEGGLEEIQGFETVG	3803
Qy	3358	-----RPMFAHFIGRLRHQESIEGATATLRCELSKAAP---VEMRK-GRESLRDG	3403
Db	3804	DTASSKTDGTORGAP-BFVELLRSCVTTEKQOALLKCV-KGEPRPKIKWTKKEKEVEMSA	3861
Qy	3404	DRHSRLROGAVCELOICGLAVADAGYSVCGSEERTSA-----TLTVKALP-----AK	3451
Db	3862	RVRAEHKDDGTLLTLPDNTQADAGYRCEAENEYGSAMTEGPIIVTLEGAPKIDGEAPD	3921
Qy	3452	FTEGLRNEEAVGATAMLCELS-KVAP-VEMKKGPNENLRDGDYILR---QESTRCELQI	3507
Db	3922	FLOFPKPAVVTVGETAVLEGKISGKPKSVKTKNGEELKPSDRVKIENLDDGTQ-RLTV	3980
Qy	3508	CGLAMADAGEYLCVCGQE---RTSATLITRALPAR---FIEDVKNQEAEGATAVLQC	3559
Db	3981	TNAKLDMDDEYRCEASNEFGVMSDVTLTVKE-PAQVAPGFFKELSAIQVKETETAKFEC	4039
Qy	3560	ELNSAAP-VEMRKGSETLRDGD--YSLRODGTKELOIRGLAMADTGEY---SCVCGQ	3612
Db	4040	KVSGTKPDVXKFKDGTPLKEDKRVHPFESTDDGTQ-RLVIEDSKTDDQGNTRIIEVSDAGV	4098
Qy	3613	ERTSAMLTVRALP---IKFTEGRLNEATEGATAVLRCEL-SKMAPVEMWKGHETLRDGD	3668
Db	4099	ANSKVPITV--VSEETLKIKKGLTDVNVTOGTKILLSVEVEGPKTKVKYKGTETVTSQ	4156
Qy	3669	RHSLRQ-DGARCELQIRGLVAEDAGEYLCMCXKE---RTSAMLTVR-----AMPSKFI	3717
Db	4157	TTKIVQVTESEYKLEIESAEMSDTGAYRVVLSTDSFVSSESTVTVTKAAEKISLPS-FK	4215
Qy	3718	EGLRNEATEGDTATLWCEL-SKAAPVEMRKGHETLRDGDHSLRQDGSCELOIRGLAV	3776
Db	4216	KGLADQVPGKTPVLVEIEGPKDKVKYKNGEIDKGVKEDL---GNGKYLTIPIPFQE	4273
Qy	3777	VDAGEYSCV---CGQERTSATLTVRALPARFIEDVKNQEAEGATAVLQCEL-SKAAPV	3831
Db	4274	KDVEYSVTAANEAGEIESAKVNVSAKP-EIVSGLVPTTVKQGETATFNVKVPKGVKV	4332
Qy	3832	BWRKGSSETL-----RGGDRYSRLRODGTCELOIHGLSVADTGEYSCV---CGQERTS	3880
Db	4333	KWYKNGKEIPDAKTKDNGDGSYS-----LEIPNAQVEDAADYKVVVNSNDAGDADS	4393
Qy	3881	ATLTVRAP-----QPVFREPLQSLQAEBSSTATLOCELSSEPTATVYWSKGLQLQANG	3933
Db	4384	AALTVKLADGDKVKKEIIVSGLIPTTVKQGETATFNVKVPKGVKQVKYKNGKEI-PNA	4442
Qy	3934	RREPRLOGCTAELVLQLOREDTGEYTCCTGSOA-----TSATLTVTAAPVRFLELQHOE	3989
Db	4443	KAKDNGDG-SYSLEIPNAQLDDTDADYKVVVNSNDAGDADSAAALTVKLPGLIAVKGLEDAE	4501
Qy	3990	VDEGTAHLCCCELRSAGASVEMRKGSLQLPFCAKVMVQDQ-AAAEELLVRGVEQEDAGDY	4048
Db	4502	VPRGKKAVALQVETNKKPKKEIKWYKNGKEIIPSDKAQFGSDGDNKFPQVLPDAGDDDAEY	4561
Qy	4049	-----TCDTGH-TQSMASLSVRP--RPKFKTRLOSLEQETGDI-ARLCCQLSDAESAVV	4100
Db	4562	KWLTDEBNTADSSCALTVKLPKPEPKI---IKGLEDDQVVISGIPKLEIETSGSPKTV	4618
Qy	4101	QWLKEGVELHAGPKVEMRSGQATELLIHQLEA---KDTGEYACVTGQKTA--SLRVT	4155
Db	4619	KWYKNGELPQAAKTIKQIKIDNKKVLEIPSSVEEDTGYKVEVANEAGSANSOKIT	4678
Qy	4156	-EPEVTIVRGIVDAEVTADEVFEFSCEVSRAGATGVQWCLQGLFLQSNVEVTEAVRDGRI	4214
Db	4679	VEPKITFLKPKDQSIIEGENAEFSVEINTKPRI-VKWKYKNGQBIKPNRSFIIQBQDITK	4737



Db 6261 TBEYGFKKLNTASLPTTPDRGPFKEBVTGHYLTLSWIFTKAPRPYQVTVIIEIRLPE 6320  
QY 6263 ---ALQETGSGPPVT---GTSE-APAVPP-----RVQPLLHE 6293  
Db 6321 KQMSLLEYNIPPEVCVRNLELGSYQVRVRAENIYG:SDPSASPRLMAPQPVDR 6380  
QY 6294 ---GPEBEPEALARAQE---W-----TVPIRMGE- 6316  
Db 6381 RTNKVILPLDPAEAKALDMRYSEQYACAPWFSPGVYKRYCAENDTLTILVNVSGFPDPD 6440  
QY 6317 AAWPGAGTCELLWDVHSHVRETTQRT---YTYOALDTHTA----- 6354  
Db 6441 IKWKFRG-----WDI-----DTSPTSCKYVTVGGSETTLAITGFSKENVQYQCPAK 6499  
QY 6355 -----RPSMQVTIEDVQAQGTGAQ---FEALIEGDPQPSVTWYKDSV 6395  
Db 6490 NDYGAQAINMVDLATRNFICPLVN---KTFSSAQPMRMDVRVDGEPFPELKKMKWR 6545  
QY 6396 QIADVSTRLSQOQEGT-TYSLVLRHVASKADAGVYTCIAQNTGCVLCKAELLVGLGDNEPD 6454  
Db 6546 PIVESSRIKFVQDGYLCSLIIINDEPMWRDSGLYSCVAVNDAGQATSTCTVTVEABGDYND 6605  
QY 6455 SEKQSHR-----RKLHSFYVEKEEIGRGVGFVKRVOHKGKILCAAKFPILRSTRQA 6509  
Db 6606 VELPRRVRTIESRRVRELYEISE-----KDEKL--AABGAPFRVKEKATG 6648  
QY 6510 -----YREDILALSHPLVTGLLDQFETRTKLILILELCSSEELLDRLYR 6555  
Db 6649 REFLAQLREIDALAKRHVDIHSNLPBGIVQHRVLRDEKALVVD--NANSTIDGLSS 6706  
QY 6556 -----KGVTEAEVKYVIOQLVEGLHLSHGVLHDIKPNSILMVHPAREDIK 6604  
Db 6707 LAHPGVETAEKGVNRETVCVRFVQQLLALXGMDHLRAHLDRPETILL---QDDKLK 6763  
QY 6605 ICDFGAQNTPAELQFOSGSPFVSPBEIIQONPVSEASDTWANGVTSYLSLTCSGSPA 6664  
Db 6764 LADFGQARLLRGLITGEIKGSPFVSPBEIVRSYPLTLATDMWSTGVLTVLLTGLSPFH 6823  
QY 6665 GSDRATILNVLEGRWSSWAHLSEDAKDFIKATIQAPQAPSPAQCLSHWPF--- 6721  
Db 6824 GDNDETLANVDSQFD--SSPL-GNFSYDAGDFVXKLITEIPVSLRTVDEALDHPWIND 6881  
QY 6722 -LKSMFAEBAHFINTKQLFELARSQRSLMSYKSLVMRS---IPELLRGP----- 6770  
Db 6882 KLKTEB-----LSADTLREFYQHKWLE-----RRVFQQTSPSQILLEAILGPATAQAQ 6930  
QY 6771 PDSPLGVARHLR-----DTGSSSSSSSDNE- 6799  
Db 6931 QNAPVAPEGRRPABIYDYLRIQPKXPPPTVYVYVQPRKEHPFIDFEGQIDGDAFDPE 6990  
QY 6800 ---LAPFARAKSLPSP-----VTHS-----PPLLHPRGFL 6826  
Db 6991 GTGFEGRHPQPOIPPOQRPNOAHDSSRHHQQPHQOQOPQIPVDQYGRPLVDPR-YL 7049  
QY 6827 RPSASLPESAERSTEAPAPSPAGAPGAPPAQCVPRHVSIRLSLFHQAGSPERGA 6886  
Db 7050 NDSHRPSSLDAPPFYVDKYNPVHFDKYCRPMA-----PONLEKGLIIPQDKGETPSSH- 7104  
QY 6887 LAPGSRHRPARRHLLKGGVIAGALPG-----LRPELMEHVRVLEBEAAR-----E 6931  
Db 7105 -KKEKCHVATPIL-----ASPGGQOQOKIPMRIRGERREIEEIANRILSDIS 7155  
QY 6932 EQATLLAKAPSPETALRLPASGTHLAPGSHSHLEHDSPTTPRSPACGEAQRLPSPAG 6991  
Db 7156 EEGSIAGSLASLED-FEIP-----KDFQVEASEPSTPTLTPE----- 7191  
QY 6992 GABIRDMGHPOGSKQIPSGCGHPTAQPERSPDPSWGPAPFCHPKQGSAPQEGCSHPH 7051  
Db 7192 -VTIRE-----TIPKTPPTSPOKSPVP----- 7214  
QY 7052 AVAPCPFGSPGSGKEAPLIVSPFLGQOPAPPAPAKAS-----PPLDSKM---G 7099  
Db 7215 -----OPQGLIIPAKVYSDSILLAGLPAADKKVLEDA 7246

QY 7100 PGDISLP-GRP--KPGPCSPGSAQSSQVSSLRVSGSQVTEPSPSLDAEGWTQAE 7156  
Db 7247 ENDPSPVPGAPLFLGLEHSGDLTIDTTSASGL--IKVTSFAINLSNPK-----SPRRS 7298  
QY 7157 DLSDSPTLQPOEQATMRKFSLGGGGYAGVAGYGTFAFGDAGGMLGOGPMWARIWA 7216  
Db 7299 TPGTKSPVLUSPQEHSMVEVLIAKRG-----KPGFLPPGEL----- 7335  
QY 7217 VSQSEEEQEBEAREAESQSEQBARAESPLQVARSPPVEVGRAPTRSSPEPTWEDIGQ 7276  
Db 7336 ---AEDIDDEDAFMDRKKQ----- 7352  
QY 7277 VSLVQIRLDSGDAEAD-----TISL-DISEVDPD--YLNLSDL-----YDIKY 7317  
Db 7353 ---VKPDHGDENDFKDEKERLEKDKNRRVTNLDLDDKYRPSAFYKDDSFHGHGXDIDA 7409  
QY 7318 LPTE--PMI----FRKVPKSAQPEPPSPMAEBELAEFFPEPTWMPGELG-PHAGLEITEE 7370  
Db 7410 TPWDSHVQIGPDYILYAARGAFAFNSVRVRYRELFGMGAPTVK-QGFLGVNRDITVRER 7468  
QY 7371 SEDVDALLAAAVGRKRWSSPSRSIHPFPGRHLPDDEPAELGLRVRKVASVHEIRLK 7430  
Db 7469 RRYTD-ILRETTQGLEPKSHEQSTAL-----LQAFSATAIERIKADIKVTPC-- 7516  
QY 7431 GRPEGLEKGPPrKKPCLASFRLSGLKSWDRAPTELRLELSDETVVLGQSVTLACQVSAQ 7490  
Db 7517 -----ATKXNDGTG-----APIFTARLDVLRKNQPAIFECVNASP 7555  
QY 7491 AAQATWSKDGAPLESRRVLISATLKNFOLLTLVVVAEDLGVYTVCSVNALGTVTGGV 7550  
Db 7556 APKVTWDFQKILESNDRVITEQD--NNVARLIINHAAPYDLGBYVCTVTAINEYGTDKSSCR 7614  
QY 7551 LRKAERSSSPCDIGEVYAD-GVLLVWVKPVSYP-----VTYIVQCSL-----EGSW 7599  
Db 7615 LISGETSPRGRPE-AELSSDTEIFQWEAPE--GPTYLEGITYRLEVRVAGNHDGDPW 7671  
QY 7600 TTLASDIFDCYLTSLKSRGTYTFRACVSKAGMPYS-----SPSEQV--- 7644  
Db 7672 ITVSEKIDDESIVKHLSPGLIYQFRVTAQNGFGLPLSLSSRIVQTHGKGAQKLQIDVL 7731  
QY 7645 -----LLGPGSHLASEESQGRSA-----QPLSTKTFAPQTQ 7677  
Db 7732 KSEIRLVNVMPQKSTNQLGISE-ESEEDSEARTANEDMKSNLQOTDDPTGRFOIGGL 7790  
QY 7678 IQGRFVSWRQOCWEKAS-GRA-LAAKIIIPVHPKDKTAVLREYEALKGLRHPHQAHAAY 7735  
Db 7791 KFKGRFSVIDAVDSTTEGHACAVKI--RHPSE--AISEYSLRDGQHENVQRLIAAF 7846  
QY 7736 LSPRHVLILELCSGPPELLCLAEASYSSEVKYLMQMLSATQVYHNOHILHLDRSE 7795  
Db 7847 NNSNFLYLLSERLY-EDVFSRFVNDYTEEQVALTMRQVTSALHFLHFKGIAHLDVNPH 7905  
QY 7796 NMIITEYN--LLKVVDLGNASLSQEKVLPSPDKFQVLET--NAPEL-LEGQAVPOTDI 7850  
Db 7906 NIMFQSKRSWVKLVDFGRQAKVS-GAVKPD-----FDYKVASPBEHIFETPVTVQSDM 7959  
QY 7851 WAIGVTAFIMLSAEYVSVSESGARDLQRLKGLVRLSRCYAGL-----SGGAVAFRLSTL 7905  
Db 7960 WGMGVVTFCLLAGHPFTSE--YDREBEIKENVINV-KCDPNLIPVNASQECLSFATWAL 8016  
QY 7906 CAQWAGPCASSCLOCPWLTGEGPACSRPAPVTFPBARL 7945  
Db 8017 KKSVPVRMTDEALSHKFLSDPSDMVRRRESIKYSASRLR 8056

## RESULT 14

Q8NH8 PRELIMINARY; PRT; 646 AA.  
ID Q8NH8  
AC Q8NH8;  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Obscurin (Fragment).  
GN OBCSN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Young P., Ehler E., Gautel M.;  
RT "Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere  
RT assembly";  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ314898; CAC85746.1; -.  
DR InterPro; IPR003961; FN.III.  
DR InterPro; IPR008957; FN.III-like.  
DR InterPro; IPR007110; IG.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003598; IG\_c2.  
DR Pfam; PF00041; fn3; 1.  
DR Pfam; PF00047; IG; 4.  
DR SMART; SM00060; FN3; 1.  
DR SMART; SM00409; IG; 6.  
DR SMART; SM00408; IGc2; 4.  
DR PROSITE; PS0835; IG LIKE; 4.  
KW Immunoglobulin domain.  
FT NON\_TER 1  
FT NON\_TER 646  
SQ SEQUENCE 646 AA; 70258 MW; B04F5B00E4323D68 CRC64;  
Query Match 7.5%; Score 3113; DB 4; Length 646;  
Best Local Similarity 93.5%; Pred. No. 8.3e-159;  
Matches 605; Conservative 16; Mismatches 24; Indels 2; Gaps 1;  
QY 421 NLLRLPKRTAVRGDTAMFCVELAVPGVPHVLRNQEEVAGGRVAISAEGRHTLTIS 480  
DB 1 NLLRLPKRTAVRGDTAMFCVELAVPGVPHVLRNQEEVAGGRVAISAEGRHTLTIS 60  
QY 481 QCLEDVGQVAFMAGDCOTSTFCVSAPRKPLPPDPVPPVKARMESSVILSWSPPPHGE 540  
DB 61 QCLEDVGQVAFMAGDCOTSTFCVSAPRKPLPPDPVPPVKARMESSVILSWSPPPHGE 120  
QY 541 RPTVIDGYLVEKKLGTVTWIRCHAEWVATPELTADVABEGNFQFRVSALNSFGQSPY 600  
DB 121 RPTVIDGYLVEKKLGTVTWIRCHAEWVATPELTADVABEGNFQFRVSALNSFGQSPY 180  
QY 601 LEFPCTVHLAKLAVRTPKAVQAVEGGEVTFSDLTVASGENFLDQALKASSVYEIH 660  
DB 181 LEFPCTVHLAKLAVRTPKAVQAVEGGEVTFSDLTVASGENFLDQALKASSVYEIH 240  
QY 661 CDRTHTLTIREVPASLHGAQLKPVANGIESSIRMEVRAAPGLTANKPPAAAAAREVLARL 720  
DB 241 CDRTHTLTIREVPASLHGAQLKPVANGIESSIRMEVRAAPGLTANKPPAAAAAREVLARL 300  
QY 721 HEEAQLLAELSDQAAAVTWLKDGRTLSPGPKYEQASGRVILVRDVARDDAGLYECVS 780  
DB 301 HEEAQLLAELSDQAAAVTWLKDGRTLSPGPKYEQASGRVILVRDVARDDAGLYECVS 360  
QY 781 RGRRTAYQLSVQGLARFLHDKMAGSCVDVAVAGGPAQPECETSEAHVHVHVYKDGMELGHS 840  
DB 361 RGRRTAYQLSVQGLARFLHDKMAGSCVDVAVAGGPAQPECETSEAHVHVHVYKDGMELGHS 420  
QY 841 GERFLQEDVGTGHRHLVAATVTRODEGTYSYCVGSDSDVDFRLRVSEPKVVFAXEQALARKL 900  
DB 421 GERFLQEDVGTGHRHLVAATVTRODEGTYSYCVGSDSDVDFRLRVSEPKVVFAXEQALARKL 480  
QY 901 QAEAGASATLSCVAQAQTEVTWYKDGKLSKSKVCWEATGCTFRLRVQAGQADAGEY 960  
DB 481 QAEVGSATLSCVAQAQTEVTWYKDGKLSKSKVCWEATGCTFRLRVQAGQADAGEY 540  
QY 961 SCEAGGQRLSFHLDVSEPKVFAKDQVAHSEVQAEAGANATLSCVAQAQTEVTWYKDGK 1020  
DB 541 SCEAGGQRLSFHLDVSEPKVFAKDQVAHSEVQAEAGANATLSCVAQAQTEVTWYKDGK 598

QY 1021 KLSSSLKVVHVEAKGCRRLRVVQAGKTDAGDYSCEARGQVSRFLHI 1067  
DB 599 KLSSSSKVCWEATGCTFRLRVVQAGQADAGEYSCGAGGQRLSFHLDV 645  
RESULT 15  
Q8NH7  
ID Q8NH7 PRELIMINARY; PRT; 551 AA.  
AC Q8NH7;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Obscurin (Fragment).  
GN OBCSN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Young P., Ehler E., Gautel M.;  
RT "Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere  
RT assembly";  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ314900; CAC85747.1; -.  
DR InterPro; IPR003599; IG.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003598; IG\_c2.  
DR Pfam; PF00047; IG; 6.  
DR SMART; SM00409; IG; 6.  
DR SMART; SM00408; IGc2; 6.  
DR PROSITE; PS0835; IG LIKE; 6.  
KW Immunoglobulin domain.  
FT NON\_TER 1  
FT NON\_TER 551  
SQ SEQUENCE 551 AA; 59655 MW; 97D8C06CF57F822F CRC64;  
Query Match 6.8%; Score 2793; DB 4; Length 551;  
Best Local Similarity 99.5%; Pred. No. 1.1e-141;  
Matches 548; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 978 PKVFAKDQVAHSEVQAEAGANATLSCVAQAQTEVTWYKDGKLSKSKVHVEAKGCR 1037  
DB 1 PKVFAKDQVAHSEVQAEAGANATLSCVAQAQTEVTWYKDGKLSKSKVHVEAKGCR 60  
QY 1038 RLWVQAGKTDAGDYSCEARGQVSRFLHIPEKMFVAKQSVHNEVQAEAGASAMLSCE 1097  
DB 61 RLWVQAGKTDAGDYSCEARGQVSRFLHIPEKMFVAKQSVHNEVQAEAGASAMLSCE 120  
QY 1096 VAQAQTEVTWYKDGKLSKSKVCWEATGCTFRLRVQAGQADAGEYSCGAGGQVSRFL 1157  
DB 121 VAQAQTEVTWYKDGKLSKSKVCWEATGCTFRLRVQAGQADAGEYSCGAGGQVSRFL 180  
QY 1158 HITEPKGVFAKQSVHNEVQAEAGANATLSCVAQAQTEVTWYKDGKLSKSKVMEVK 1217  
DB 181 HITEPKGVFAKQSVHNEVQAEAGANATLSCVAQAQTEVTWYKDGKLSKSKVMEVK 240  
QY 1218 GCTRLRVQVQAGKADAGEYSCGAGGQVSRFLHIPEKMFVAKQSVHNEVQAEAGASAT 1277  
DB 241 GCTRLRVQVQAGKADAGEYSCGAGGQVSRFLHIPEKMFVAKQSVHNEVQAEAGASAT 300  
QY 1278 LSCEVAQAQTEVTWYKDGKLSKSKVCWEATGCTFRLRVQAGQADAGEYSCGAGGQRL 1337  
DB 301 LSCEVAQAQTEVTWYKDGKLSKSKVCWEATGCTFRLRVQAGQADAGEYSCGAGGQRL 360  
QY 1338 SFHLDVSEPKVFAKQSVHNEVQAEAGANATLSCVAQAQTEVTWYKDGKLSKSKSVR 1397  
DB 361 SFHLDVSEPKVFAKQSVHNEVQAEAGANATLSCVAQAQTEVTWYKDGKLSKSKSVR 420  
QY 1398 MEAVGCTRLRVQVQAGQADAGEYSCGAGGQVSRFLHIPEKMFVAKQSVHNEVQAEAG 1457  
DB 421 MEAVGCTRLRVQVQAGQADAGEYSCGAGGQVSRFLHIPEKMFVAKQSVHNEVQAEAG 480

Qy	1458	ASTTISCEVAQAQTEVVMYKDGKKLSFSSKVRMEAVGCTRRLLVVQQAGQADAGEYSCEAG	1517
Db	481	ASTTISCEVAQAQTEVVMYKDGKKLSFSSKVRMEAVGCTRRLLVVQQAGQAVAGEYSCEAG	540
Qy	1518	SORLSFHLHVA	1528
Db	541	SORLSFHLHVA	551

Search completed: September 13, 2004, 11:35:51  
Job time : 636 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:07:04 ; Search time 2831.1 Seconds  
(without alignments)  
10242.013 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_23150\_24120  
Perfect score: 971  
Sequence: 1 cagcggcgccggcgctggcgcg.....ggccgctaaaaagtctaga 971

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues  
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_esti.\*

9: gb\_esti.\*

10: gb\_est2.\*

11: gb\_hic.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estom.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gssprt.\*

21: em\_gssfun.\*

22: em\_gssmam.\*

23: em\_gssmus.\*

24: em\_gsspro.\*

25: em\_gssrod.\*

26: em\_gssphg.\*

27: em\_gssvri.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	713.4	73.5	715	12	BM985339
C 2	672.2	69.2	731	12	EG818748
C 3	626.4	64.5	989	13	EX435097
C 4	480	49.4	480	13	EX280322

C 5	399.8	41.2	640	14	CD774776
C 6	377.4	38.9	622	14	CF176441
C 7	339	34.9	768	10	BF134040
C 8	331	34.1	479	10	BF651426
C 9	313.4	32.3	522	13	EQ554402
C 10	311.4	32.1	493	10	BE110731
C 11	247.4	28.5	494	14	CF177813
C 12	243.4	25.1	428	9	AI503993
C 13	238.2	24.5	408	10	BF414891
C 14	235.2	24.2	418	13	EX636992
C 15	231.6	23.9	445	13	BY588786
C 16	227.6	23.4	411	12	BI186010
C 17	214.8	22.1	494	14	CB713368
C 18	209.8	21.6	417	13	BY437537
C 19	209.6	21.6	379	10	BE754373
C 20	205.8	21.2	364	10	BF073878
C 21	200.8	20.7	370	9	AI716505
C 22	200.4	20.6	3241	11	AK035543
C 23	199.4	20.5	485	14	CB727068
C 24	196.8	20.3	401	13	BY438883
C 25	192	19.8	359	12	BM124337
C 26	190.6	19.6	355	12	BM123571
C 27	188.4	19.4	345	10	AW784908
C 28	175.6	18.1	357	10	BE667930
C 29	164.4	16.9	621	12	EG792248
C 30	156.4	16.1	1030	12	BM552361
C 31	150	15.4	397	10	AW479627
C 32	147.2	15.2	276	9	AI763601
C 33	147	15.1	410	9	AI716969
C 34	144.2	14.9	569	10	BE757743
C 35	144	14.8	274	13	EQ943127
C 36	138.4	14.3	749	14	CB494554
C 37	138	14.2	304	12	BM031895
C 38	137.8	14.2	629	12	BI429497
C 39	135.8	14.0	553	12	BM185681
C 40	133.8	13.8	683	10	AW862431
C 41	128.4	13.2	995	13	EQ722567
C 42	128.2	13.2	187	10	AW435707
C 43	123	12.7	392	12	BM186727
C 44	122.4	12.6	750	28	BZ173917
C 45	120.8	12.4	567	12	BI345924

ALIGNMENTS

RESULT 1

BM985339/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

BM985339 715 bp mRNA linear EST 20-FEB-2003  
UI-CF-EC1-acg-j-11-0-UI-s1 UI-CF-EC1 Homo sapiens cDNA clone  
UI-CF-EC1-acg-j-11-0-UI-3', mRNA sequence.

BM985339  
UI-CF-EC1-acg-j-11-0-UI-3', mRNA sequence.

BM985339.1 GI:19611739

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 715)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

889546

Contact: McCray, PB

McCray Lab

University of Iowa

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com) or from Open Biosystems  
 (www.openbiosystems.com).  
 Seq primer: M13 FORWARD  
 PolyA=yes.

## FEATURES

source

Location/Qualifiers  
 1..715  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-CF-ECL-acg-j-11-0-UI"  
 /tissue\_type="Lung"  
 /dev\_stage="Adult and Fetal"  
 /lab\_hosts="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-CF-ECL"  
 /notes="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-CF-ECL is a normalized cDNA library containing the  
 following tissue(s): Normal lung from adult and from fetal  
 day 64, day 87, week 19 and week 42. The library was  
 constructed according to Bonaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 adaptor, digested with Not I, and cloned directionally  
 into pT73-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dT)18 tail. The sequence tag for this library is  
 AAGTGCCTTAC.  
 TAG TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371  
 and 380-383  
 TAG\_LIB=UI-CF-ECL  
 TAG\_SEQ=AAGTGCCTTAC"

## ORIGIN

Query Match 73.5%; Score 713.4; DB 12; Length 715;  
 Best Local Similarity 99.9%; Pred. No. 1.3e-81;  
 Matches 714; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 210 ACTCAGAAATCCGAGGTGAAGACTACTGTGGCAGATGTTGAGTGCACCCAGTACCTGC 269  
 DB 715 ACTCAGAAATCCGAGGTGAAGACTACTGTGGCAGATGTTGAGTGCACCCAGTACCTGC 656

QY 270 ACACACGACATCCCTGCACCTGCAGCTGAGGTCCGAGAACATGATCATCCGCAATACA 329  
 DB 655 ACACACGACATCCCTGCACCTGCAGCTGAGGTCCGAGAACATGATCATCCGCAATACA 596

QY 330 ACTGTCTCAAGGTGCTGGACCTCGGCAATGACAGAGCTTCAGCCAGGAGAGGTGCTGC 389  
 DB 595 ACCTGTCTCAAGGTGCTGGACCTCGGCAATGACAGAGCTTCAGCCAGGAGAGGTGCTGC 536

QY 390 CTTACAGACATGTTCAAGGACTACTAGACACATGCTCCAGAGTCTCTGGAGGCGCAGG 449  
 DB 535 CTTACAGACATGTTCAAGGACTACTAGACACATGCTCCAGAGTCTCTGGAGGCGCAGG 476

QY 450 GGGCTGTTCACAGACAGACATCTGGGCCATCGGTGTGACAGCTTTCATCATGCTGAGCG 509  
 DB 475 GGGCTGTTCACAGACAGACATCTGGGCCATCGGTGTGACAGCTTTCATCATGCTGAGCG 416

QY 510 CCGAGTACCCGTTGAGCAGCAGGAGTGCACCGACCTGCAGAGAGACTGCCAAGGGCG 569  
 DB 415 CCGAGTACCCGTTGAGCAGCAGGAGTGCACCGACCTGCAGAGAGACTGCCAAGGGCG 356

QY 570 TGGTCCGGCTGAGCCGCTGTACCGGGGCTGTCCCGGGGCGCGCTGGGCTTTCCTGCGCA 629  
 DB 355 TGGTCCGGCTGAGCCGCTGTACCGGGGCTGTCCCGGGGCGCGCTGGGCTTTCCTGCGCA 296

QY 630 GCATCTGTGCGGCCAGCCCTCGGGCGCGCCCTGCGCGTCCAGCTGCCTGAGTGCCTCGT 689

Db 235 GCATCTGTGCGCCAGCCCTGGGGCGGGCCCTCGCGGTCCAGCTGCTGAGTGCCCGT 236  
 QY 690 GGCTAACACAGAGAGAGGGCCCGCTGTTTCGGCGCCCGCGCGCTGACCTTCCCTACCGCGC 749  
 Db 235 GGCTAACACAGAGAGAGGGCCCGCTGTTTCGGCGCCCGCGCGCTGACCTTCCCTACCGCGC 176  
 QY 750 GGCTGCGCGCTCTTCTGTCGGCAATCGCAGAGAGAGAGCGCGCTGCTCTACAGAGGACCA 809  
 Db 175 GGCTGCGCGCTCTTCTGTCGGCAATCGCAGAGAGAGAGCGCGCTGCTCTACAGAGGACCA 116  
 QY 810 ACCTGGGCCAGGTGCTGAGGGTTCGCCCGCGGCACACCCCTTGGTCTCCCGCTGGGGGT 869  
 Db 115 ACCTGGGCCAGGTGCTGAGGGTTCGCCCGCGGCACACCCCTTGGTCTCCCGCTGGGGGT 56  
 QY 870 CGCTGCAGACCGCGCAATAAAGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 924  
 Db 55 CGCTGCAGACCGCGCAATAAAGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1

RESULT 2  
 BG818748 731 bp mRNA linear EST 22-MAY-2001  
 LOCUS 602779074F2 NCI\_CGAP\_Brn67 Homo sapiens cDNA clone IMAGE:4914402  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BG818748  
 VERSION BG818748.1 GI:14166335  
 KEYWORDS EST  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 731)  
 NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-re@mail.nih.gov  
 Tissue Procurement: David N. Louis, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM10821 row: d column: 19  
 High quality sequence stop: 693.

FEATURES  
 source  
 1..731  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4914402"  
 /tissue\_type="anaplastic oligodendroglioma with ip/19q  
 loss"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NCI\_CGAP\_Brn67"  
 /note="Organ: Brain; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 2.3 kb. Constructed by Life  
 Technologies. Note: this is a NCI\_CGAP Library."

## ORIGIN

Query Match 69.2%; Score 672.2; DB 12; Length 731;  
 Best Local Similarity 97.9%; Pred. No. 2.2e-76;  
 Matches 702; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

QY 220 CGAGGTGAAGGACTACCTGTGGCAGATGTTGAGTGCACCCAGTACCTGCACAAACAGCA 279  
 Db 1 CGAGGTGAAGGACTACCTGTGGCAGATGTTGAGTGCACCCAGTACCTGCACAAACAGCA 60

QY 280 CATCTCTGACCTGCAGCTGAGGTCCGAGACATGATCATCCAGGAGACATGATCATCCAGGAGTCA 339  
 Db 61 CATCTCTGACCTGCAGCTGAGGTCCGAGGAGACATGATCATCCAGGAGTCA 120







Db 580 GTTATAGACCTGGGAAATGCCAGAGTCTCGACCAAGAGAGGTCCTCCAGCCCTGAGAAC 521  
QY 401 TTCAGAGGACTACCTAGAGACCAATGGCTCCAGAGCTCTCGAGGGCGGAGGGGCTGTTC 460  
Db 520 TTCAGAGGACTACCTAGAGACCAATGGCTCCAGAGCTCTCGAGGGCGGAGGGGCTGTTC 461  
QY 461 CAGACAGACATCTGGGCCATCGGTGTGACAGCTTCATCATCTGAGCGCCGAGTACCG 520  
Db 460 CAGACAGACATCTGGGCTATTGGTGTACAGCTTCATATGCTGAGTGGGAGTACCA 401  
QY 521 GTGAGCAGCGAGGTGACGAGCACTGTGAGAGAGACTGCGCAAGGGGCTGGTCCGGCTG 580  
Db 400 GAGAGCAGCGAGGGGACTTCGCGACCTGCGAGAAAGGCTTGGCGCAAGGAGACTCATTCGGTTG 341  
QY 581 AGCCGCTCTACGCGGGGCTGTCCGGGGCGCGCTGCGGCTTCCTGCGCAGCACTCTGTC 640  
Db 340 AGTCGCTCTATGAGGATTATCAGAGAGCGGTAGCTTCTTCGAGAGTTCATTATGT 281  
QY 641 GCCAGCCCTGGGGCGGCTGTGCGCGTCCAGCTGCTGAGTCCCGTGGCTTAACAGAG 700  
Db 280 GCTCAACCTGGGGCGGCTGTGCGCGTCCAGCTGCTGAGTCCCGTGGCTTAACAGAG 221  
QY 701 GAGGGCGGCTGTTCGCGCGCGCGCTGAGCTTCCCTAGCGCGGCTGCGGCTG 760  
Db 220 GAGGGCGGCTGTTCGCGCGCGCGCTGAGCTTCCCTAGCGCGGCTGCGGCTG 161  
QY 761 TTCGTGCGCAATCGCGAGAGAGAGCGCGCTGTGTACAGAGGCGCAACCTCGCCCGAG 820  
Db 160 TTGTGCGCGAGCGCGAGAGCGGTGCGCTCTCTACAGAGCATACCTGCGCCAG 101  
QY 821 GTGCGCTAGGGTGCCTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 880  
Db 100 GTGCGCTAGG---TCCAGCTCTACAGAGCAAGATGTGCCATGCCATGCGGGGACACC 44  
QY 881 CGCCAAATAAAACGCCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 923  
Db 43 CGCTAATAAAGATGCCAAACGCGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1

RESULT 6  
CF176441/c  
LOCUS 800530 MARC 3P1G Sus scrofa cDNA 3', mRNA linear EST 28-JUL-2003  
DEFINITION CF176441  
VERSION CF176441.1 GI:33288217  
KEYWORDS EST.  
SOURCE Sus scrofa (pig)  
ORGANISM Sus scrofa

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
1 (bases 1 to 822)  
Smith,R.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J.,  
Wise,T.A., Nonneman,D.J., Wray,J.E. and Keele,J.W.  
A second set of porcine ESTs from a pooled-tissue normalized  
library

JOURNAL Unpublished (2003)  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called with phred v0.020425.c and  
trimmed with the aid of the trim\_alt option. Vector identified with  
cross\_match v0.990329.  
Seq primer: SRG8004 row: M column: 6  
Location/Qualifiers

FEATURES  
source 1..622  
/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/tissue\_type="pooled"

/lab\_host="DH10B"  
/clone\_lib="MARC 3P1G"  
/notes="Vector: pcDNA3.1; Site: 1: EcoRI; Site: 2: NotI;  
Library made with RNA pooled from multiple tissues  
including brain, liver, muscle, placenta/endometrium,  
ovary, testes, and bone marrow."

ORIGIN  
Query Match 38.9%; Score 377.4; DB 14; Length 622;  
Best Local Similarity 81.9%; Pred. No. 6.8e-39;  
Matches 435; Conservative 0; Mismatches 96; Indels 0; Gaps 0;  
QY 299 AGTCCGAGAAATGATCATCACCGAATACACCTGCTCAAGTCTCGTGGACCTGGGCAAT 358  
Db 621 AGTCCGAGAAATGATCATCACCGAATACACCTGCTCAAGTCTCGTGGAAAT 562  
QY 359 GCACAGAGCTCAGCAGAGAGAGTGTGCGCTCAGACAAAGTTCAGAGTACCTACCTAGAG 418  
Db 561 GCCCGAGATTCTCACTCAGGAGAGTGTCTGCGCTCAGAGCGGTTCAGAGTATGTGGAG 502  
QY 419 ACCATGCGCTCCAGAGCTCTGGAGGGCCAGGGGCTGTTCACAGACAGACATCTCGGCG 478  
Db 501 ACTATGGCTCCAGAGCTCATAGAGGCGAGGTGTCTCCACAGACTGATATATGGGCC 442  
QY 479 ATCGGTGTGACAGCTTCATCATGTAGCGCGGAGTACCCGCTGAGCGAGGCTGCA 538  
Db 441 ATAGGTGTCACTGCCCTTCATCATGTAGTGTGTAGTACCCAGTGAAGCGAGGCGAGC 382  
QY 539 CGCGACTGACAGAGAGTGTGCGAGGGGCTGTGCGCTGAGCGCTGCTACGCGGG 598  
Db 381 AGAGATTTCAGAAAGGCTTGGCGAAGGGGCTCATCCAGATGAGCGGATGCTACGCGGG 322  
QY 599 CTGTCCGGGGGGCGCGCTTCTGCGCAGACCTCTGTGCGCGCCAGCCCTGGGGCGCGG 658  
Db 321 CTCTCCGGGGGTGCGCTGCGCTTCTTCCGAGCACCTGTGTGCACACCCCTGGGGCGGA 262  
QY 659 CCCTGCGCTGACAGTGTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 718  
Db 261 CCATGTGCTCCAACTGTGTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 202  
QY 719 CGGCGCGCGCGCTGACCTTCCCTACCGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 778  
Db 201 CACCCAGACACTGTGACCTTCCACAGACAGCTGTGCGCACACCCCTGGGGCGGA 142  
QY 779 AAGAGAGCGCGCTGCTGTACAGAGGACAACTTGGCGCCAGCGTGGCTGA 829  
Db 141 AAGAGGCGGCACTGTGTACAAAAGACAACTGACCCAAAGTGCATGA 91

RESULT 7  
BF134040  
LOCUS 601778492F1 NCI\_CGAP\_Lu30 Mus musculus cDNA clone IMAGE:4006702 5',  
DEFINITION mRNA sequence.  
ACCESSION BF134040  
VERSION BF134040.1 GI:10973080  
KEYWORDS Mus musculus (house mouse)  
SOURCE EST.  
ORGANISM Mus musculus

REFERENCE NIH-MGC http://mgi.nci.nih.gov/  
AUTHORS 1 (bases 1 to 768)  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: NCI-CCAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov  
 Plate: LIA9238 row: 9 column: 23  
 High quality sequence stop: 736.  
 FEATURES  
 source  
 1. 768  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="Czech 11"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4006702"  
 /tissue\_type="tumor, metastatic to mammary"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Lu30"  
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; transgenic model Wnt-1, expression driven by MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies.  
 ; Investigator providing samples: Gilbert Smith, NIH"

## ORIGIN

Query Match 34.9%; Score 339; DB 10; Length 768;  
 Best Local Similarity 79.6%; Pred. No. 4.8e-34;  
 Matches 450; Conservative 0; Mismatches 110; Indels 5; Gaps 4;

QY 2 AGCGGGGGGGCTGGCGCCCAAGATCATCCCTTACCACCCCAAGGACAGACAGCAGTG 61  
 DB 179 AGTGGGGGGCCCTAGCTGCTAAGATCGTTCCTTACCACCTGAGGACAGACAGCTGTA 239  
 QY 62 CTGCGCGAATACGAGGCGCTCAAGGGCTGGCGCCACCCGACCTGGCCCGAGCTGCAGCA 121  
 DB 239 CTAAGAGAATACGAGGCACTTAAGAGACTGACACCCACATCTGGCCCAACTCCATGCC 299  
 QY 122 GCCTACCTCAGCCCGGACCTGGTCTCATCTTGGAGCTGTCTCTGGGCCCGAGCTG 181  
 DB 299 GCCTACCTCAGTCCCGGACCTGGTCTCATCTTGGAGCTGTCTCTGGGCCCGAGCTG 358  
 QY 182 CTCCTCCCTGGCGGAGGGCCCTCTCTACATCCAGATCCGAGGTGAAGACTACCTGTGG 241  
 DB 359 CTACCTCTTTGGGGGAGGAGAACTCTACTCAGAGTCTGATGTGAAGGACTACCTGTGG 418  
 QY 242 CAGATGTGAGTGCACCCAGCTACTGACACACAGACATCTGCACTGGACCTGGACCTGAGG 301  
 DB 419 CAGATGTGAGTGCACCCAGTACTTGCATGCCACACATCTGCACTGGACCTGGACCTGAGG 478  
 QY 302 TCCGAGAACATGATCATCACCAGATACAACCTGCTCAAGTCTGTGGACCTGGGCAATGCA 361  
 DB 479 TCCGAGAACATGATGTTCACTGAGTACAACTGCTTAAGTTATAGACTGGGAAATGCC 538  
 QY 362 CAGAGCTCAGCCAGGAGAGTCTGCTCCCTCAGACAGTTCAGAGACTACCTAGAGACC 421  
 DB 539 CAGAGTCTCGACCAAGAGAGGTCCTCAGACCTCTGAGAACTTCAAGAGACTACCTGGAGACC 598  
 QY 422 ATGGCTCCAGAGCTCTCTGGAGGCGCAGGGGCTGTTCACAGACAGACATCTGGGCGATC 481  
 DB 599 ATGGCTCCAGAACTCTGGAAGGTCAACCGGCTGTTCACAGACAGACA--TTGGGCTAT 656  
 QY 482 GGTGTGACAGCTTTCATCTGAGCGCCGAGTACCGGCTGAGCAGGAGGTGACGC 541  
 DB 657 GGTGTAAACAG--CTTCAATTATCTGAGTGGGAGTA--CCAGAGACAGGAGGGGAGCTCGT 714  
 QY 542 GACCTGCAGAGAGGACTCGCAAGG 566  
 DB 715 CA-CTGGAGAAAGGCTTCGGCAGGG 738

## RESULT 8

BF651426  
 LOCUS BF651426 479 bp mRNA linear EST 25-APR-2001  
 DEFINITION 274209 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.  
 ACCESSION BF651426  
 VERSION BF651426.1 GI:11916556  
 KEYWORDS EST.  
 SOURCE Bos taurus (cow)

## ORGANISM

Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
 Bovidae; Bovinae; Bos.

## REFERENCE

1 (bases 1 to 479)  
 Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,  
 Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C.,  
 Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A.,  
 Chitko-McKown, C.G., Peitea, G., Holt, I., Karamycheva, S., Liang, F.,  
 Quackenbush, J. and Keefe, J.W.

## TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA  
 libraries and construction of a gene index for cattle

## JOURNAL

## MEDLINE

## PubMed

## COMMENT

Genome Res. 11 (4), 626-630 (2001)

Contact: Smith RPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@mail.marc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred

vo.980904.e. Vector identified by cross\_match with the -minscore 18

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCACGTCACGACG

Plate: 63 row: L column: 10

Seq primer: ATTAGGTGACACTATAG.

## FEATURES

## source

1..479

/organism="Bos taurus"

/mol\_type="mRNA"

/db\_xref="taxon:9913"

/tissue\_type="pooled"

/lab\_host="DH10B"

/clone\_lib="MARC 3BOV"

/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;

Library made from pooled tissue from marrow, alveolar

macrophage, ovary, fetal semitendinosus muscle, and fetal

longissimus muscle."

## ORIGIN

Query Match 34.1%; Score 331; DB 10; Length 479;  
 Best Local Similarity 82.0%; Pred. No. 5.8e-33;  
 Matches 393; Conservative 0; Mismatches 85; Indels 1; Gaps 1;

QY 323 GAATACAACTGCTCAAGTCTGTGGACCTGGGCAATGCACAGAGCCTCAGCCAGAGAG 382  
 DB 2 GAGTATACCTGCTC-AGGTATCGACTTCGGAACGCCAGAGTTTGGCCAGAGAGG 60  
 QY 383 GTGCTGCCCTCAGACAAAGTTCAAGACTACCTAGAGACCATGGTCCAGAGCTCTCGAG 442  
 DB 61 GTCTCTGCCCTCAGAGAGATTCAAGGACTACATGGAGACCATGGCTCCGAGCTCCTGGAG 120  
 QY 443 GSCCAGGGGCTGTTCACACAGACATCTGGGCACTGGGTGACAGCCTTCATCATG 502  
 DB 121 GSCCAGGGGCTGTTCGCGAGACTGACATCTGGGCCATAGGTGTACAGCCTTCATCATG 180  
 QY 503 CTGAGCGCCGAGTACCCCGTGAGCAGCGAGGTGCACGCGACCTGTCAGAGAGAGCTGGCC 562  
 DB 181 TTGAGCGCTGACTACCCGCTGAACCGTGAAGGAAACGCGAGACACGACAGAGGCTGGCC 240  
 QY 563 AAGGGGCTGCTCCGCTGAGCGCTGTACGCGGGGCTGTCCGGGGGCGCGTGGCCCTTC 622  
 DB 241 AAGGGGCTCATCCAGCTGAGCGCTGTCTATGCCGGGCTGTCTGGGGGCTGTGGCCCTTC 300  
 QY 623 CTGCGCAGCACTCTCTGGGCCAGCCCTCTGGGCGCGGCTTCGCGGTCCAGTGTGCTGAG 682  
 DB 301 CTCCGAAGCACACTGTGTGTGCACCCCTGGTCCGGCGGTGTGCATCCAGCTGCTGCGAG 360  
 QY 683 TGCCTGCTGCTAACAGAGAGGGGCGCCGCTTTCGCGGCGCGCGCTTCCCTTCCCT 742

```

Db      361 AGCCCGTGGTGAACGAGAGAGGGCCACAGCTCCACGCCCGGGGGTGAACCTCCCG 420
QY      743 ACCGCGCGCGTCCGCGTTCCTGCGCAATCGCAGAGAGACGCGGCTGCTGTACAA 801
Db      421 ACCACTGCTGCGCGTCTTTGTGCGAGAGCGCGAAGAGCGCGGCACTGCTGTACAA 479

RESULT 9
BQ554402/c
LOCUS   H4027H08-3 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
DEFINITION H4027H08 3', mRNA sequence.
ACCESSION BQ554402
VERSION   BQ554402.1 GI:21455290
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE
AUTHORS   VanBuren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G.,
          Martin,P.R., Stagg,C.A., Bassey,U., Aiba,K., Hamatani,T.,
          Kargul,G.J., Luo,A.G., Keiso,J., Hide,W. and Ko,M.S.H.
          Assembly, verification, and initial annotation of NIA 7.4K mouse
          cDNA clone set
          Genome Res. 12 (12), 1999-2003 (2002)
          2354164
          12466305
          Other ESTs: H4027H08-5
          Contact: Yong Qian
          Laboratory of Genetics
          National Institute on Aging/National Institutes of Health
          333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
          Email: cdna@igsun.grc.nia.nih.gov
          This clone set has been freely distributed to the community. Please
          visit http://lgsun.grc.nia.nih.gov/cDNA/NIA\_7.4k.html for details.
          Plate: H4027 row: H column: 08
          Seq primer: -21M13 Forward
          High quality sequence stop: 522
          POLYA=Yes
FEATURES             Location/Qualifiers
     source           1..522
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="nia:EST:H4027H08-3"
                     /db_xref="taxon:10090"
                     /clone="H4027H08"
                     /sex="mixed"
                     /dev_stage="mixed"
                     /lab_host="DH10B"
                     /clone_lib="NIA Mouse 7.4K cDNA Clone Set"
                     /note=Vector; pSPORT1; Site_1: Sali; Site_2: NotI; This
                     clone is among a rearranged set of 7,407 clones from more
                     than 20 cDNA libraries."
ORIGIN
Query Match      32.3%; Score 313.4; DB 13; Length 522;
Best Local Similarity 80.9%; Pred. No. 9.7e-31;
Matches 365; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY      381 AGGTGTCCTCCAGACAGTTCAAGCACTACCTAGAGACATGCTCCAGAGCTCTCTGG 440
Db      522 AGGTCCCGACGCCCTGAGAACTTCAAGACTACCTGGAGACCATGGCTCCAGAACTCCTGG 463
QY      441 AGGCCACAGGGGCGTTCCACAGACACACATCTGGGCCATCGGTGTGCACAGCCTTCATCA 500
Db      462 AAGTCAAGGGCGTGTCCACAGACACACATTTGGGCTATTTGGTTHAACAGCTTCATTA 403
QY      501 TGCTGAGCGCGAGTACCCGGTGAAGCAGCGAGGTTGCACGCGACTGTCAGAGAGGACTGC 560
Db      402 TGCTGAGTGGCGAGTACCCAGAGACGAGCGAGGGAATCTCGCGACCTTCGAGAAAGGCTTGC 343

```

```

QY      561 GCAAGGGGTGTTCCGGCTGAGCCGCTGTACGCGGGGCTGTCCGGGGCGCGCTGGCT 620
Db      342 GCAAGGGACTCATTCGGTTGAGTGTGCTATATGAGGATTATCAGAGAGGCGGTAGCT 283
QY      621 TCTTGGCGAGCACTGTGTCGCCGCCAGCCCTGGGGCGGGCCCTGGCGGTCCAGCTGCTGC 680
Db      282 TCTTGCAAGATTCAATTATGTCTCAACCTGGGGCGCGCTTCACCTGCTTGC 223
QY      681 AGTGCCCGTGGTTAAACAGAGAGGGGCCCGCTGTTCGGCGGCCCGCCCGTGACCTTCC 740
Db      222 AATGCGGATGGTGCACAGAGAGGGGCCACCGGCTCCCGGCCACGCCGTAACCTTCC 163
QY      741 CTACCGCGGGCTGCGCGCTCTTCGTGCGCAATCGGAGAGAGAGCGCGGCTGTGTGACA 800
Db      162 CCACCGTGGGTTGCGCGCTTTGTGCGGAGCGGAGAAACGGGCTGCGCTCTCTACA 103
QY      801 AGAGGCACACACCTGGCCCGAGGTGCGCTGAGG 831
Db      102 AGNAGCATAACTGGCCAGGTGCGCTGAGG 72

BE110731 493 bp mRNA linear EST 13-JUN-2000
UI-R-BJ1-avd-g-08-0-UI.s1 UI-R-BJ1 Rattus norvegicus cDNA clone
UI-R-BJ1-avd-g-08-0-UI 3', mRNA sequence.
BE110731
LOCUS BE110731.1 GI:8502877
DEFINITION Rattus norvegicus (Norway rat)
ACCESSION BE110731
VERSION BE110731
KEYWORDS EST.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 493)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
        discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
        Coordinated Laboratory for Computational Genomics
        University of Iowa
        375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
        Tel: 319 335 8250
        Fax: 319 335 9565
        Email: bento-soares@uiowa.edu
        The sequence contained an oligo-dT track that was present in the
        oligonucleotide that was used to prime the synthesis of first
        strand cDNA and therefore this may represent a bonafide poly A
        tail. The sequence tag present in the cDNA between the NotI site
        and the oligo-dT track served to identify it as a clone from the
        normalized heart library cDNA library Preparation: M.B. Soares Lab
        Clone distribution: clones will be available through Research
        Genetics (www.resgen.com)
        Seq primer: M13 Forward
        POLYA=Yes.
FEATURES             Location/Qualifiers
     source           1..493
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BJ1-avd-g-08-0-UI"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-BJ1"
                     /note=Vector; pYT3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1
                     library is a subtracted library derived from the following
                     tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
                     canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,
                     AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.

```

For a detailed description of the library from which this clone was derived, please visit our web site at [rategest.eng.uiowa.edu](http://rategest.eng.uiowa.edu). The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)  
 TAG TISSUE=heart  
 TAG LIB=UT-R-BJ1  
 TAG\_SEQ=ACAAC"

## ORIGIN

Query Match 32.1%; Score 311.4; DB 10; Length 493;  
 Best Local Similarity 82.4%; Pred. No. 1.8e-30;  
 Matches 357; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 407 GACTACCTAGAGACCATGGCTCCAGAGCTCTCGAGGGCCAGGGGGTGTTCACAGACA 466  
 Db |||||  
 QY 493 GACTACCTAGAGACCATGGCTCCAGACTTCTGAGGGCCAGGGGGGTTCACAGACA 434  
 Db |||||  
 QY 467 GACTCTGGGCGATCGGTGTGACAGCCCTTCATCATGTGAGCGCCGAGTACCCGGTGAGC 526  
 Db |||||  
 QY 433 GACATCTGGGCTATTGGTGTAAACAGCCCTTCATTATGCTGAGTGGCGAGTACCCAGTGAGC 374  
 Db |||||  
 QY 527 AGCAGAGGTGACGCGACCTGCGAGAGGACTCGCAAGGGGTGTCGGGTGAGCCGCG 586  
 Db |||||  
 QY 373 AGCAGAGGGGACTCGGACCTGCGAAGAGGCTGCGCAAGGACTCAATTGAGTCGC 314  
 Db |||||  
 QY 587 TGCTACCGGGGGTGTCCGGGGCGCGCTTCTCGCGAGCACTCTGTGCGCCGAG 646  
 Db |||||  
 QY 313 TGCTATGACGAGTATCAGGGGTGCGGTAGACCTTCTCGCAGAGTTCATTGCGCTCG 254  
 Db |||||  
 QY 647 CCTTGGGGCGGCTCGGCTGAGTGCCTGAGTGCCTGAGTGCCTGAGTGCCTGAGTGCCT 706  
 Db |||||  
 QY 253 CCTTGGGGTGGCGGCTGCTTCCACTGCTGAGTGGGGTGGCTGAGGAGGAGGCG 194  
 Db |||||  
 QY 707 CCGGCTGTTCCGGGGCGGCGGCTGACCTTCCCTACCGCGGGGTGCGCGCTTCCTG 766  
 Db |||||  
 QY 193 CCGGCTGTTCCGGGGCGGCGGCTGACCTTCCCTACCGCGGGGTGCGCGCTTCCTG 134  
 Db |||||  
 QY 767 CGCAATCGGAGAGAGACGGCGCTCTGTACAGAGGACAACTGCGCGAGTGCCTG 826  
 Db |||||  
 QY 133 CGCAGGCGGAGAGAGCGGGCGCTACTCTACAGAGACAACTGCGCTCAGGTGCGC 74  
 Db |||||  
 QY 827 TGAGGGTGGCGCC 839  
 Db |||||  
 QY 73 TGAGGGCGGCGCC 61  
 Db |||||

RESULT 11  
 CF177813  
 LOCUS 494 bp mRNA linear EST 28-JUL-2003  
 DEFINITION 805332 MARC 3P1G Sus scrofa cDNA 5', mRNA sequence.  
 ACCESSION CF177813  
 VERSION CF177813.1 GI:33289589  
 KEYWORDS EST.  
 SOURCE Sus scrofa (pig)  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 1 (bases 1 to 494)  
 AUTHOR Smith, T.P.L., Freking, B.A., Ford, J.J., Vallet, J.L., Fox, J.,  
 Wise, T.A., Noneman, D.J., Wray, J.E. and Keele, J.W.  
 TITLE A second set of porcine ESTs from a pooled-tissue normalized library  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Smith, TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@mail.marc.usda.gov  
 Single pass sequencing. Bases called with phred v0.020425.c and  
 trimmed with the aid of the trim\_alt option. Vector identified with  
 cross\_match v0.990329.

Plate: SRG8004 row: M column: 6  
 Seq primer: GTAATAGCACTCACTATAGG.

## FEATURES

## source

1..494  
 Location/Qualifiers  
 /organism="Sus scrofa"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9823"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /clone\_lib="MARC 3P1G"  
 /note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;  
 Library made with RNA pooled from multiple tissues  
 including brain, liver, muscle, placenta/endometrium,  
 ovary, testes, and bone marrow."

## ORIGIN

Query Match 25.5%; Score 247.4; DB 14; Length 494;  
 Best Local Similarity 82.3%; Pred. No. 2.4e-22;  
 Matches 284; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 1 CAGCGGGCGGCGTGGCGCCCAAGATCATCCCTACACCCCAAGACAGACAGCAGT 60  
 Db |||||  
 QY 150 CATTGGGGCGGCTGGCTGCTGCAATATCGTGGGGGCGCCCTGAGAACAGGACTGCCGT 209  
 Db |||||  
 QY 61 GCTCGCGAATACGAGGCGCTCAAGGGCTGCGCCACCCGACCTGCCAGCTGCACGC 120  
 Db |||||  
 QY 210 GCTTCGATAATATGATGCTTCAAGGGCTTGGCCACCTCACCTGCACATTTGCGAGC 269  
 Db |||||  
 QY 121 AGCTTACCTCAGCCCCCGGCACTGGTGTCTCATCTTGAGGTGTCTTGGGCCGAGCT 180  
 Db |||||  
 QY 270 TGCTTACCTCATCCCGGCGACCTGGTCTCATCTTGGAGTTGTCTTGGCGCTGATCT 329  
 Db |||||  
 QY 181 GCTCCCTGCTGCGCGAGAGGGCTCTCTACTAGATCCGAGTCCGAGTGAAGACTACCTGTG 240  
 Db |||||  
 QY 330 GCTCCCTGCTGCGCGAGAGGGCTCTCTACTATATATATATATATGAAAGACTATCTGTG 389  
 Db |||||  
 QY 241 GCAGATGTTGAGTGCACCCAGTACCTGCACCAACAGCACATCTGCACCTGCACCTGAG 300  
 Db |||||  
 QY 390 GCAGATGCTGATGGCACTTACCTGCAGCCGAGCTATCTGCATCTAGACCTCAT 449  
 Db |||||  
 QY 301 GTCGAGAACATGATCATCCGCAATACACCTGCTCAAGTCT 345  
 Db |||||  
 QY 450 GTCCGAAAACATGATGTGCACCGAGTACACCTGCTCAAGATCAT 494  
 Db |||||

## RESULT 12

## AI503993/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## UNPUBLISHED (1999)

## CONTACT

## WASHINGTON UNIVERSITY SCHOOL OF MEDICINE

## 4444 FOREST PARK PARKWAY, BOX 8501, ST. LOUIS, MO 63108, USA

## TEL: 314 286 1800

## FAX: 314 286 1810

## EMAIL: mouseest@wustl.edu

## THIS CLONE IS AVAILABLE ROYALTY-FREE THROUGH LLNL; CONTACT THE

## IMAGE CONSORTIUM (INFO@IMAGE.LLNL.GOV) FOR FURTHER INFORMATION.

MG1:565403  
This clone was previously sequenced on the 5' end only, this new data is from the 3' end  
High quality sequence stop: 427.  
Location/Qualifiers  
1..428  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1001187"  
/tissue\_type="diaphragm"  
/dev\_stages="adult"  
/lab\_host="SOLR (kanamycin resistant)"  
/clone\_lib="Stratagene mouse diaphragm (#937303)"  
/note="Organ: diaphragm; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally from mRNA prepared from diaphragm muscle. Primer: Oligo dt. Average insert size: 1.5 Kb. Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGCAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

FEATURES

source

ORIGIN

Query Match 25.1%; Score 243.4; DB 9; Length 428;  
Best Local Similarity 80.1%; Pred. No. 8.1e-22;  
Matches 286; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
QY 475 GCCATCGGTGACAGCCTTCATCATGCTGAGCGCCAGTACCGGTGACGAGGG 534  
DB 428 GGCATTGGTGTAAAGCCTTCATTATGCTGAGTGGCGAGTACCCAGAGACAGCGGG 369  
QY 535 TGCAGCGACCTGCGAGAGAGACTCGCAAGGGGTGTCTCGGTGAGCCGCTCTACGC 594  
DB 368 GACTCGGACCTGCAGAAAGGCTTGGCAAGGACTATTGGTTAGTCTGCTATGC 309  
QY 595 GGGGCTGTCGGGGCGCGCTTCTCTCGGAGCACTGTGCGCCAGCCCTGGG 654  
DB 308 AGGATTATCAGAGAGCGGTAGCCCTCTCGAGAGTTCATTATGTCTCAACCTGGG 249  
QY 655 CGGCGCTCGCGTCCAGTGCCTGCTGAGTCCGCTGCTAAACAGAGAGCGCCGCGTG 714  
DB 248 CGGCCGCTGGCTTCCACCTGCTTGAATGCGGTGCTGACAGAGAGGCGCCACCGG 189  
QY 715 TTCGGGGCGCGCGCGCTTCCCTACCGCGCGGTGCGGTCTTCGTGCGCAATCG 774  
DB 188 CTCGGGGCGCGCGCGCTTCCCTACCGCGCGGTGCGGTCTTCGTGCGCAATCG 129  
QY 775 CGAGAAGAGAGCGCGCTGCTGTACAAGAGGACAACTCGCCAGGTGCGCTGAGG 831  
DB 128 CGAGAAGCGCGTGGCTCTCTCTACAGAGCATATCTGCCAGGTGCGCTGAGG 72

RESULT 13  
BF414891/c  
LOCUS  
DEFINITION  
UI-R-BJ2-bot-c-07-0-UI-s1 UI-R-BJ2 Rattus norvegicus cDNA clone  
ACCESSION  
BF414891  
VERSION  
BF414891.1 GI:11402880  
KEYWORDS  
EST.  
SOURCE  
Rattus norvegicus (Norway rat)  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 408)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL  
Genome Res. 6 (9), 791-806 (1996)  
MEDLINE  
9704477  
PUBMED  
889548  
COMMENT  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics

University of Iowa  
375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized heart library cDNA library Preparation: M.B. Soares Lab Genomics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=Yes.  
Location/Qualifiers  
1..408  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-BJ2-bot-c-07-0-UI"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-BJ2"  
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ2 library is a subtracted library derived from the following tissues: heart, atrium at 15 dpc, ventricle at 15.5 dpc, atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15 dpc, AV canal at 15 dpc. For a detailed description of the library from which this clone was derived, please visit our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)  
TAG\_TISSUE=heart  
TAG\_LIB=UI-R-BJ2  
TAG\_SEQ=ACAAAC"

FEATURES

source

Query Match 24.5%; Score 238.2; DB 10; Length 408;  
Best Local Similarity 80.4%; Pred. No. 3.8e-21;  
Matches 279; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
QY 493 CTTTCATCTGCTGAGCGCGAGTACCCGGTGTGACGAGGGTGACGAGCTGCAGAG 552  
DB 408 CTTTCATTATGCTGAGTGGCGAGTACCCAGTGTGACGAGGGGACTCGCGACTGCAGAA 349  
QY 553 AGGACTGCGCAAGGGGCTGTCGGCTGAGCGCTGCTACGCGGGGCTGTCGGGGGCGC 612  
DB 348 AGCCCTGCGCAGGGGACTCATTCATTGAGTCTCTATGAGATTATCAGGGGGTGC 289  
QY 613 CGTGGCTTCTCTGCGCAGCACTCTGTGCGCCAGCCCTGCGGCGCGCTGCGCGTCCAG 672  
DB 288 GGTAGCTTCTCTGACAGTTCATTGTGCGCTCGGCCCTGGGTGCGCGCTTCCAC 229  
QY 673 CTGCGTGTGAGTCCCGTGTGCTTAACAGAGAGGCGCGGCTGTTCGCGCGCGCGCGCT 732  
DB 228 CTGCTTGTGAGTGTGCGGTGTGCTGACGAGGAGGCGCGGCTTCCGCGCGCGCGCT 169  
QY 733 GAGCTTCCCTTACCGCGCGCTGCGCTCTTCGTGCGCAATCGCGAGAGAGAGCGCGCT 792  
DB 168 GACCTTCCCGCGCGGATTCGGTGCCTTGTGCGGAGCGCGGAGAGCGCGCGCT 109  
QY 793 GCTGTACAAGAGGACAACTGTGCGCCAGGTGCGGTGAGGGTGCCTCC 839  
DB 108 ACTCTACAAGAGGACAACTGTGCGCTCAGGTGCGGTGAGGCCCGCGCC 62

RESULT 14  
EX636992/c  
LOCUS  
DEFINITION  
BX636992 pBluescript Lion Mus musculus cDNA clone LiONp462F07402  
3', mRNA sequence.

EX636992 418 bp mRNA linear EST 12-AUG-2003  
BX636992 pBluescript Lion Mus musculus cDNA clone LiONp462F07402  
3', mRNA sequence.

BX636992  
 BX636992.1 GI:33616867  
 EST.  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 418)  
 Henrich, J., Hermans, J., Kranz, H., Loebbert, R., Schlueter, T.,  
 Schuette, D., Weindel, M., Heil, O., Ebert, L., Neubert, P., Peters, M.,  
 Radelof, U., Schneider, D. and Korn, B.  
 Mouse ArrayTAG cDNA (LION)  
 Unpublished (2003)  
 Contact: Ina Rolfs  
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
 Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany  
 RZPD; LIONp462F07402.  
 RZPDLIB;  
 Mouse ArrayTAG cDNA (LION)  
 http://www.rzpd.de/cgi-bin/products/showLib.pl.cgi/?response?libNo=4  
 62 Contact: Ina Rolfs  
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH  
 Heubnerweg 6, D-14059 Berlin, Germany  
 Tel: +49 30 32639 101  
 Fax: +49 30 32639 111  
 www.rzpd.de  
 This clone is available royalty-free from RZPD:  
 contact RZPD (clone@rzpd.de) for further information. Seq primer:  
 RP: CAGGAACACGATGAC.  
 Location/Qualifiers  
 1..418  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="LIONp462F07402"  
 /lab\_host="DH10B"  
 /clone\_lib="pbluescript Lion"

Query Match 24.2%; Score 235.2; DB 13; Length 418;  
 Best Local Similarity 80.2%; Pred. No. 9e-21;  
 Matches 276; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 498 ACAGCCTTCATCTGAGCGCCGAGTACCCGCTGAGCAGCGAGGTGACGCGACTG 547  
 |||||  
 Db 417 ACAGCCTTCATCTGAGTGGCGAGTACCCGAGAGCAGCGAGGCGACTGCGCGCTG 358  
 |||||  
 QY 548 CAGAGGAGCTGCCAAGGGCTGGTCCGCTGAGCGCTGCTACGCGGGGCTGTCGGG 607  
 |||||  
 Db 357 CAGAAAGCTTCCGCAAGGACTCATTCGGTTGAGTCCGCTATGAGGATATACGGA 298  
 |||||  
 QY 608 GCGCGCTGGCTTCCTCTCGCAGCAGCTCTGCGCCCGCAGCGCTGGCGCGCGCTGCGG 667  
 |||||  
 Db 297 GGAGCGGTAGCTTCTCTCGACAGTTCATTATGCTCAACCTGGCGCGCGCTGGGCT 238  
 |||||  
 QY 668 TCCAGCTGCTGAGTCCCTGCTGCTACAGAGAGGCGCGCTGCTGCGCGCGCGG 727  
 |||||  
 Db 237 TCCAGCTGCTGCAATGGGATGCTGACAGAGGCGCGCGCTGCTGCGCGCGCGG 178  
 |||||  
 QY 728 CCGGTGACTTCCCTACCGCGCGCTGCGCGCTCTCTGCTGGCAATCCGAGAGAGAGCG 787  
 |||||  
 Db 177 CCGGTGACTTCCCTACCGCGCGCTGCGCGCTCTCTGCTGGCAATCCGAGAGAGAGCG 118  
 |||||  
 QY 788 GCGCTGCTGACAGAGCACAACCTGCCCGAGTGGCTGAGG 831  
 |||||  
 Db 117 GCGTCTCTCAAGAAGCATAATCTGCCCGAGTGGCTGAGG 74  
 |||||

RESULT 15  
 BY588786 445 bp mRNA linear EST 15-DEC-2002  
 LOCUS  
 BY588786 RIKEN full-length enriched, adult inner ear Mus musculus  
 DEFINITION  
 cDNA clone F93003104 3', mRNA sequence.

BY588786  
 BY588786.1 GI:26923968  
 EST.  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 445)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamataka, I.,  
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
 Schorbach, C., Gojibori, T., Baldarelli, R., Hill, D. P., Buit, C.,  
 Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,  
 Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Bruscia, V.,  
 Chothia, C., Corbani, L. E., Cousins, S., Della, E., Dragani, T. A.,  
 Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,  
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
 Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,  
 Kawaji, H., Kawasaki, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,  
 Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,  
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
 Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G.,  
 Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,  
 Ravasi, T., Reed, J. C., Reid, J., Ring, B. Z., Ringwald, M.,  
 Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,  
 Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,  
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
 Wells, C., Wilming, L. G., Wyshaw-Boris, A., Yanagisawa, M., Yang, I.,  
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,  
 Rogers, J., Birney, E. and Hayashizaki, Y.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 12466851  
 22354683  
 COMMENT  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suihito-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
 Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,  
 Miyazaki, A., Murata, K., Nakamura, M., Nomura, K., Numazaki, R.,  
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and  
 Hayashizaki, Y. Direct Submission  
 Computational Analysis of Full-length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Tissues were provided by Kirk W. Beisel ( Boys Town National  
 Research Hospital 555 North 30th Street Omaha, NE 68131 USA ) whose



assistance we gratefully acknowledge.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

```
FEATURES
  source
    1..445
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="F930003104"
      /tissue_type="inner ear"
      /dev_stage="adult"
      /clone_lib="RIKEN full-length enriched, adult inner ear"
```

ORIGIN

```
Query Match      23.9%; Score 231.6; DB 13; Length 445;
Best Local Similarity 79.7%; Pred. No. 2.5e-20;
Matches 298; Conservative 0; Mismatches 74; Indels 2; Gaps 2;

Qy 459 CACAGACAGACATCGGCCATCGGTGTGACAGCCTTCATCATCTGACGCGCGAGTACC 518
Db 1 CACAGACAGACATTGGGCTATTGGTGTAAACGCTTCATTATGCTGAGTGGCGAGTACC 60

Qy 519 CGGTGAGCAGCAGGGTGCACGCCACCTGCAGAGAGGACTCGCAAGGGGCTGGTCCGGC 578
Db 61 CAGAGAGCAGCAGGG-ACTCGTGACCTTCAGAAAGGCTTGCGCAAGGGACTCATTCGTT 119

Qy 579 TGAGCCGCTGTACCGGGGCTGTCCGGGGCGCCGTCGGCTTCCTCCGAGCAGACTCTGT 638
Db 120 TGAGTCGCTGTATGACAGATTATCAGGAGAGCGGTAGCCTTCCTGCAGATTTCATTAT 179

Qy 639 GCGCCCGACCTGGGGCGGCCCTGCGCTCCAGCTGCCTGCGCGCCGTGGCTAACAG 698
Db 180 TTGCTCAACCTGGGGCGGCCCTGCGCTCCAGCTGCCTGCAATGCGGTGGCTGACAG 239

Qy 699 AGGAGGGCCCGGCTGTTCGGGCGCCCGCCGTGACCTTCCTTACC-CCGCGGCTGGCC 757
Db 240 AGGAGGGTCCACCGGCTCCCGGCCCAAGCCGTAACCTTCCCGCCGCGGCTGGCC 299

Qy 758 GTCTTCGTCCGAATCGCGAGAGAGCGCGCTGTGTACAGAGGACACACCTGGCC 817
Db 300 GCCFTTGTGCGCGAGCGGAGAGCGGCTGCGCTCCTCTACAAAGACATTAACCTGGCC 359

Qy 818 CAGGTGCGCTGAGG 831
Db 360 CAGGTGCGCTGAGG 373
```

Search completed: September 19, 2004, 22:07:15  
Job time : 2840.1 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: September 19, 2004, 15:09:24 ; Search time 34.5752 Seconds  
(without alignments)  
6179.453 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_16862\_17246  
Perfect score: 385  
Sequence: 1 ccgagggagatcttgacac.....acctgcagcactggagcgc 385

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	38.8	10.1	723	4	US-09-252-991A-5281
C 2	38.8	10.1	744	4	US-09-252-991A-5345
C 3	38.8	10.1	954	4	US-09-252-991A-5307
C 4	37.2	9.7	1200	4	US-09-907-754A-22
C 5	37.2	9.7	1200	4	US-09-905-125A-22
C 6	37.2	9.7	1200	4	US-09-902-775A-22
C 7	37	9.6	4080	4	US-09-016-434-1342
C 8	35	9.1	523	4	US-09-621-976-1358
C 9	34.8	9.0	486	4	US-09-252-991A-1276
C 10	34.6	9.0	2028	4	US-09-252-991A-13079
C 11	34.6	9.0	3525	4	US-09-252-991A-12708
C 12	34.6	9.0	3906	4	US-09-252-991A-13251
C 13	34.4	8.9	855	4	US-09-252-991A-9597
C 14	34.4	8.9	1080	4	US-09-252-991A-9553
C 15	34.4	8.9	1098	4	US-09-252-991A-13156
C 16	34.4	8.9	1101	4	US-09-252-991A-12790
C 17	34.4	8.9	1131	4	US-09-252-991A-9699
C 18	34.2	8.9	4403765	3	US-09-103-840A-2
C 19	34.2	8.9	4411529	3	US-09-103-840A-1
C 20	34	8.8	430	4	US-09-621-976-16656
C 21	34	8.8	1909	3	US-09-100-193-6
C 22	34	8.8	4112	1	US-08-340-203A-2
C 23	34	8.8	4112	2	US-08-452-567-2
C 24	34	8.8	4112	2	US-08-452-427-2
C 25	34	8.8	4112	3	US-09-085-407-2
C 26	34	8.8	4287	1	US-08-244-189-1
C 27	34	8.8	4287	1	US-08-306-691B-53

Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 13, Appli  
Sequence 63, Appli  
Sequence 76, Appli  
Sequence 812, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 190, Appli  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 1899, Ap  
Sequence 3, Appli  
Sequence 14, Appli

1 US-08-340-203A-1  
2 US-08-452-567-1  
3 US-08-452-427-1  
4 US-09-085-407-1  
5 US-08-943-768-1  
6 US-08-865-960-1  
7 US-09-398-858-13  
8 US-09-266-965-63  
9 US-09-266-965-76  
10 US-09-023-655-812  
11 US-08-209-747-1  
12 US-08-458-298-1  
13 US-09-495-050A-190  
14 US-08-103-840A-2  
15 US-09-103-840A-1  
16 US-09-489-039A-1899  
17 US-09-811-469-3  
18 US-08-403-852D-14

8.8 4616 1  
8.8 4616 2  
8.8 4616 3  
8.8 4616 4  
8.8 3150 3  
8.8 3150 4  
8.7 1173 4  
8.7 1425 4  
8.7 53500 4  
8.6 2407 4  
8.6 2793 1  
8.6 2793 1  
8.6 2793 1  
8.6 1935 4  
8.6 4403765 3  
8.6 4411529 3  
8.5 702 4  
8.5 83450 4  
8.4 474 2

C 28 34  
C 29 34  
C 30 34  
C 31 34  
C 32 33.8  
C 33 33.8  
C 34 33.6  
C 35 33.6  
C 36 33.6  
C 37 33.2  
C 38 33.2  
C 39 33.2  
C 40 33  
C 41 33  
C 42 33  
C 43 32.8  
C 44 32.6  
C 45 32.4

ALIGNMENTS

RESULT 1  
US-09-252-991A-5281/c  
; Sequence 5281, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 5281  
; LENGTH: 723  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5281

Query Match 10.1%; Score 38.8; DB 4; Length 723;  
Best Local Similarity 47.2%; Pred. No. 0.51;  
Matches 118; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy 25 CGTGGTCACCGCTGACTACTGCTCCCTAGGGCTGAGCAGGATGCCATCACGCTGGCGGA 84  
Db 542 CGAGATCGTCGGGACCGAGCGCTCCGGGGCGATGCCACGCGGAGCCGCGCCGCG 483  
Qy 85 AGCCAGTATGTGGAGGTCTCTGGATGACGCCACCCACCTGCGCTGGCTTGTTCGACACCA 144  
Db 482 CTTCCAGCAGGTAGTAGATGCTCGAACCCCTGGGCATAGCGCAGCGCTCCGCGCCA 423  
Qy 145 GCCACCAAGTCAGCCCTCAGCGAGGCTGGGTGTCCAGAGCTTCTTGGACAGGAG 204  
Db 422 GCTCTTGGTTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 363  
Qy 205 GCTCAAGCTGTACCTGAGTGGGGGGCGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGG 264  
Db 362 GCTCTTGCACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 303  
Qy 265 GTCTGAAGAC 274  
Db 302 GGCTGACCAC 293

RESULT 2  
US-09-252-991A-5345/c  
; Sequence 5345, Application US/09252991A

Patent No. 6551795  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 CURRENT FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 5345  
 LENGTH: 744  
 TYPE: DNA  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-5345

Query Match 10.1%; Score 38.8; DB 4; Length 744;  
 Best Local Similarity 47.2%; Pred. No. 0.52; Mismatches 132; Indels 0; Gaps 0;  
 Matches 118; Conservative 0;  
 QY 25 CGTGGTACCGCTGACTACCTGCCCTAGGCGGTGAGCAGGATGCCATCAGCTGCGGGA 84  
 Db 573 CGAGATCGTCGCGACCAAGCGCTCCGGGGGATGGCCACGCCGAGGCGCCACCGCG 514  
 QY 85 AGGCAGATGTGGAGTCTGGATGCGAGCCACCCACTGCGTGGCTTGTCCGCCCAA 144  
 Db 513 CTTCCAGCAGGTAGTAGAGTCTCGAACCCCTGGCCATAGCGAGCGCTCCGCGCCCA 454  
 QY 145 GCCACCAAGTCCAGCCCTCAGCGCGGCGGTGGTGTACCCAGCTACCTGCACAGGAG 204  
 Db 453 GCCCTGCTCGCGGCCCGCCAGCGCGCTGGGGTCCGAGGCGGTATGCAGCAGCG 394  
 QY 205 GCTCAAGTGTCACTGATGAGTGGGGCGCGCTGAGGCCCTGTAGTTCCTGGGGAGGCTGT 264  
 Db 393 GCTCTGCAACAGCGCGCGCGCGCGCTGGGCCAGGCGCGGTTCCTGCGCCAGGCGCG 334  
 QY 265 GTCTGAAGAC 274  
 Db 333 GGCTGACCAC 324

RESULT 3  
 US-09-252-991A-5307/c  
 Sequence 5307, Application US/09252991A  
 Patent No. 6551795  
 GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 CURRENT FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO 5307  
 LENGTH: 954  
 TYPE: DNA  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-5307

Query Match 10.1%; Score 38.8; DB 4; Length 954;  
 Best Local Similarity 47.2%; Pred. No. 0.55; Mismatches 118; Conservative 0; Gaps 0;  
 Matches 118; Conservative 0;  
 QY 25 CGTGGTACCGCTGACTACCTGCCCTAGGCGGTGAGCAGGATGCCATCAGCTGCGGGA 84  
 Db 820 CGAGATCGTCGCGACCAAGCGCTCCGGGGCGATGGCCACGCCGAGGCGCGCCACCGCG 761

QY 85 AGGCAGTATGTGGAGTCTCGATGCAGCCACCCACTGCGCTGGCTTGTCCGCCCAA 144  
 Db 760 CTTCCAGCAGGTAGTAGATCTCGAACCCCTGGCCATAGCGCAGCGCTCCGCGCCA 701  
 QY 145 GCCACCAAGTCCAGCCCTCAGCGCGGCGGTGGTGTACCCAGCTACCTGTCACAGGAG 204  
 Db 700 GCCCTGCTCGCGGCCCGCCAGCGCGCGCTGGGGTCCGAGCGGTATGCACACG 641  
 QY 205 GCTCAAGTGTCACTGATGAGTGGGGCGCGCTGAGGCCCTGTAGTTCCTGGGGAGGCTGT 264  
 Db 640 GCTCTGCAACAGCGCGCGCGCGCTGGGCCAGGCGCGGTTCCTGCGCCAGGCGCG 581  
 QY 265 GTCTGAAGAC 274  
 Db 580 GGCTGACCAC 571

RESULT 4  
 US-09-907-794A-22  
 Sequence 22, Application US/0907794A  
 Patent No. 6635468  
 GENERAL INFORMATION:  
 APPLICANT: Genentech, Inc.  
 APPLICANT: Ashkenazi, Avi  
 APPLICANT: Botstein, David  
 APPLICANT: Desnoyers, Luc  
 APPLICANT: Eaton, Dan L.  
 APPLICANT: Ferrara, Napoleone  
 APPLICANT: Filvaroff, Ellen  
 APPLICANT: Fong, Sherman  
 APPLICANT: Gao, Wei-Qiang  
 APPLICANT: Gerber, Hanspeter  
 APPLICANT: Gerritsen, Mary E.  
 APPLICANT: Goddard, A.  
 APPLICANT: Godowski, Paul J.  
 APPLICANT: Grimaldi, Christopher J.  
 APPLICANT: Gurney, Austin L.  
 APPLICANT: Hillan, Kenneth, J.  
 APPLICANT: Kijavini, Ivar J.  
 APPLICANT: Mather, Jennie P.  
 APPLICANT: Pan, James  
 APPLICANT: Paoni, Nicholas F.  
 APPLICANT: Roy, Margaret Ann  
 APPLICANT: Stewart, Timothy A.  
 APPLICANT: Tumas, Daniel  
 APPLICANT: Williams, P. Mickey  
 APPLICANT: Wood, William I.  
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
 TITLE OF INVENTION: Acids Encoding the Same  
 FILE REFERENCE: 10466-14  
 CURRENT APPLICATION NUMBER: US/09/907,794A  
 CURRENT FILING DATE: 2001-07-17  
 PRIOR APPLICATION NUMBER: PCT/US00/04414  
 PRIOR FILING DATE: 2000-02-22  
 PRIOR APPLICATION NUMBER: US 60/143,048  
 PRIOR FILING DATE: 1999-07-07  
 PRIOR APPLICATION NUMBER: US 60/145,698  
 PRIOR FILING DATE: 1999-07-26  
 PRIOR APPLICATION NUMBER: US 60/146,222  
 PRIOR FILING DATE: 1999-07-28  
 PRIOR APPLICATION NUMBER: PCT/US99/20594  
 PRIOR FILING DATE: 1999-09-08  
 PRIOR APPLICATION NUMBER: PCT/US99/20944  
 PRIOR FILING DATE: 1999-09-13  
 PRIOR APPLICATION NUMBER: PCT/US99/21090  
 PRIOR FILING DATE: 1999-09-15  
 PRIOR APPLICATION NUMBER: PCT/US99/21547  
 PRIOR FILING DATE: 1999-09-15  
 PRIOR APPLICATION NUMBER: PCT/US99/23089  
 PRIOR FILING DATE: 1999-10-05  
 PRIOR APPLICATION NUMBER: PCT/US99/28214  
 PRIOR FILING DATE: 1999-11-29

```

, PRIOR APPLICATION NUMBER: PCT/US99/28313
, PRIOR FILING DATE: 1999-11-30
, PRIOR APPLICATION NUMBER: PCT/US99/28564
, PRIOR FILING DATE: 1999-12-02
, PRIOR APPLICATION NUMBER: PCT/US99/28565
, PRIOR FILING DATE: 1999-12-02
, PRIOR APPLICATION NUMBER: PCT/US99/30095
, PRIOR FILING DATE: 1999-12-16
, PRIOR APPLICATION NUMBER: PCT/US99/30911
, PRIOR FILING DATE: 1999-12-20
, PRIOR APPLICATION NUMBER: PCT/US99/30999
, PRIOR FILING DATE: 1999-12-20
, PRIOR APPLICATION NUMBER: PCT/US00/00219
, PRIOR FILING DATE: 2000-01-05
, NUMBER OF SEQ ID NOS: 423
, SEQ ID NO 22
, LENGTH: 1200
, TYPE: DNA
, ORGANISM: Homo sapiens
US-09-907-794A-22

```

Query Match	9.7%; Score 37.2; DB 4; Length 1200;
Best Local Similarity	49.6%; Pred. No. 1.5;
Matches 125; Conservative	0; Mismatches 123; Indels 4; Gaps 1;
QY	122 CTGCGCTGGCTTGTCCGCACCAAGCCACCAAGCTCCAGCCCTTCACGGCAGGGCTGGGGTG 181
Db	715 GTGGGGTGGGGAGGGGAGCCAGATCCCCGAGGGAGGCCCTTGAGGGCCGCGAAGCATC 774
QY	182 TCACGAGCTTACCTGGACAGAGGCTCAAGCTGTCACTGAGTGGGGGGCCCTCAGGCC 241
Db	775 CGAGCCCCCAGCTGGGA----AGGGCAGGCGCGTGCCCCAGGGGGCGGTGGCACAGTGC 830
QY	242 CTTGAGTTCCTGGGGAGGCTGTGTCCTGAAGACGAATACAAGGCAAGGTGAGCTCTGTG 301
Db	831 CCCCTTCGCGACGGGTGGCAGGCCCTTGGAGAGAACTGAGTGTCACCCTGATCTCAGGC 890
QY	302 ATTCAGGAGCTGCTGAGTTCGTGACAGGCCCTTCGTGGAGAGCTGCAGTTCCTGTCAGAGC 361
Db	891 CACCAAGCCTCTCCGGCCCTCCAGCGGGGTCTCTGAAGCCCGCTGAAGGTGAGCGATG 950
QY	362 CACCACCTGCAG 373
Db	951 AAGGCTTTGCAG 962

RESULT 5  
US-09-905-125A-22  
Sequence 22, Application US/09905125A  
Patent No. 6664376  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Botstein, David  
APPLICANT: Desnovers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Flvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Garber, Hanspeter  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard A.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth, J.  
APPLICANT: Kijavlin, Ivar J.  
APPLICANT: Mather, Jennie P.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.

```

; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreated and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Secreated and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,125A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 22
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-905-125A-22

```

Query Match	9.7%; Score 37.2; DB 4; Length 1200;
Best Local Similarity	49.6%; Pred. No. 1.5;
Matches 125; Conservative	0; Mismatches 123; Indels 4; Gaps 1;
Qy	122 CTGGCTGGGTTGTCCGCACCAAGCCACCAAGTCCAGCCCTCACGCGAGGGCTGGGTG 181
Db	715 CTGGGTGGCGGAGGGGAGCCAGATCCCGAGGGAGACCCCTGAGGGCCGCGAAGCATC 774
Qy	182 TCACCAGCTACCTGGACAGAGGCTCAAGCTGCACCTGAGTGGGGGGCCGTGAGGCC 241
Db	775 CGAGCCCCCAGCTGGGA-----AGGGGAGGCCCGGTGCCCGAGGGCCGTGGCCACATGC 830
Qy	242 CCTGAGTTCCTCGGGAGGGTGTGTCTGAAGACGAATACAAAGGCAAGGCTGAGCTCTGTG 301
Db	831 CCCCTTCCCGGACGGGTGGCAGGCCCTTGGAGAGGAAGTGAAGTGCACCTTGATCTCAGGC 890
Qy	302 ATCCAGAGTGTGTAGTTCTGAGCAGGCCTTCGTGGAGGAGCTGCAGTTCCTTCACAGC 361
Db	891 CACCAAGCTCTGCGGGCTCCACGCCGGGTCTTGAAGCCCGCTGAAGAGTCAGCGACTG 950
Qy	362 CACCACCTGAG 373
Db	951 AAGGCCTTGAG 962

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-902-775A-22

Query Match          9.7%; Score 37.2; DB 4; Length 1200;
Best Local Similarity 49.6%; Pred. No. 1.5;
Matches 125; Conservative 0; Mismatches 123; Indels 4; Gaps 1

QY      122 CTGCGCTGGCTGTCCGACACCAAGCCACCAAGTCCAGGCCCTCAGCGGAGGGCTGGGTG 181
        |||||
DB      715 CTGGGGTGGCGGAGGGAGGCCAGATCCCAGGGAGACCTCTGAGGGCCGCGAAGCATC 774

QY      182 TCACCAAGCTTACCTGGACAGGAGCTTCAAGCTGTCTACCTGAGTGGGGGCCGCTGAGGCC 241
        |||||
DB      775 CGAGCCCCAGCTGGGA-----AGGGGAGGCCGGTGTGCCCCAGGGGGGGTGGCAGAGTGC 830

QY      242 CCTGAGTTCCTTGGGGAGGCTGTCTGTGAGAGCAATACAGGACAGGCTGAGCTCTGTG 301
        |||||
DB      831 CCCCTTCCGACGGGTGGCAGGCCCTCTGGAGAGAACTGAGTGTACCCCTGATCTCAGGC 899

QY      302 ATCCAGGAGCTGCTGAGTCTTGACAGAGGCCCTTCTGTGAGGAGAGCTGCAGTTCCTGCAGAGC 361
        |||||
DB      891 CACCAAGCTCTGCCGGCTCTCCAGCGGGCTCTCTGAAGCCGGCTGAAAGGTCAGCGACTG 950

QY      362 CACCACTGCAG 373
        |||||
DB      951 AAGGCTTGCAG 962
        |||||

RESULT 7
US-09-016-434-1342
; Sequence 1342, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1342:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4080 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:

```

LIBRARY: GENBANK  
CLONE: g34764  
US-09-016-434-1342

Query Match 9.6%; Score 37; DB 4; Length 4080;  
Best Local Similarity 46.4%; Pred. No. 2.2;  
Matches 121; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 95 GTGAGAGTCTCTGATGACGCCACCCACTGGCTGGCTTGTCCGACCAACAGCCACCAAG 154  
DB 263 GTGAGAGTCTCTGATGACGCCACCCACTGGCTGGCTTGTCCGACCAACAGCCACCAAG 322  
QY 155 TCCAGCCCTTACCGGAGGGCTGGGTGTACACCTTCCAGAGCTGTGACAGAGAGGTCAAGCTG 214  
DB 323 CCGTAAAGGGCTCGAAGCGCGGCGCACCGCTGCCGACAGGTTCATGAGGGCGGCGCTC 382  
QY 215 TCACCTAGTGGGGCGCGCTGAGGCGCCCTGAGTTCCTCTGGGAGGCTGTCTGAGAC 274  
DB 383 GCACCCAACTGAGCGCGGAGGAGCAACGCCAGCGCCGCGCGCGGCGGCGGAGGC 442  
QY 275 GAATACAGGCAAGCTGAGCTCTGTGATCCAGGAGCTGTGAGTTCTGAGCAGGCTTC 334  
DB 443 AACCGCACCGCGGACCCCGCGCGCAACGAGGCCCTTGGCGCGGTGAGGTGGCGGTG 502  
QY 335 GTGAGGAGCTGCAGTTCCTG 355  
DB 503 CTGTGTCTCACTGCTCCTG 523

RESULT 8

US-09-621-976-1358/c  
; Sequence 1358, Application US/09621976  
; Patent No. 8639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 1358  
; LENGTH: 523  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 176..421  
; NAME/KEY: sig\_peptide  
; LOCATION: 176..268  
; OTHER INFORMATION: Von Heijne matrix  
; OTHER INFORMATION: score 10.8999995185303  
; OTHER INFORMATION: seq LLLALSLAHLCA/FS  
US-09-621-976-1358

Query Match 9.1%; Score 35; DB 4; Length 523;  
Best Local Similarity 45.1%; Pred. No. 4.4;  
Matches 113; Conservative 1; Mismatches 131; Indels 0; Gaps 0;

QY 19 CATCTAGCTGTACCCGCTGACTACCTGCCCTAGGGGCTGAGCAGGATCCATCAGCT 78  
DB 290 CACCTGAGGTTCTAGGCTAAAGCCCGACGATAGTGGCACGTGAGAGGGCCAGGAGGA 231  
QY 79 GCGGAGGAGCCAGTATGTGGAGTCTGGATGAGCCACCCACTGGCTGGCTGTCCG 138  
DB 230 GCTGAGGCTCAGGGGTGGGGGATAGCAAGGAAGTAGAAGTGGTGTGGCATGTGC 171  
QY 139 CACCAAGCCCAACCAAGTCCAGCCCTCACGGCAGGGGTGGTGTCAACGACCTTACCTGGA 198  
DB 170 CAGTTCCACCCACCCCTTCCCTCTTAGGGAGAGGAGTGGCAGAGCAACACTGAGGC 111

QY 199 CAGGAGGCTCAAGCTGTCACTGAGTGGGGGGCGCTGAGGCCCTGAGTTCCTCTGGGA 258  
DB 110 TGCAGGACACAGACACTGCCCTATGGGGGTATGTCAGCTCTCGTTCTGTSACT 51  
QY 259 GGCTG 263  
DB 50 GGGTG 46

RESULT 9

US-09-252-991A-1276/c  
; Sequence 1276, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 1276  
; LENGTH: 486  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-1276

Query Match 9.0%; Score 34.8; DB 4; Length 486;  
Best Local Similarity 50.6%; Pred. No. 4.8;  
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 216 CACCTGAGTGGGGCGGCTGAGGCCCTGAGTTCCTCTGGGAGGCTGTCTGAAGAG 275  
DB 266 CGCTCGGGCGTAGCGGCAAGGCTTCGCGGTGAGGGAACCGCGCTTCGGGAGG 207  
QY 276 AATACAGGCAAGGCTGAGCTCTGTGATCCAGGAGCTGTGAGTTCTGAGCAGGCTTCG 335  
DB 206 ATTCCGCGCGCGCGCCATTTCGCTCCAGTTCTGTCATCAGTGGCTGTGTCGCCCG 147  
QY 336 TGGAGGAGCTGCAGTTCCTGACAGCCACCACCTCAGCACCTGGA 381  
DB 146 AGAAGCTGTTGAGGACCGGAGCTCGGCGGCTTCAGGGCCTGGA 101

RESULT 10

US-09-252-991A-13079  
; Sequence 13079, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13079  
; LENGTH: 2028  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13079

Query Match 9.0%; Score 34.6; DB 4; Length 2028;  
Best Local Similarity 52.4%; Pred. No. 7.5;  
Matches 76; Conservative 0; Mismatches 69; Indels 0; Gaps 0;





```
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9553
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9553

Query Match      8.9%; Score 34.4; DB 4; Length 1080;
Best Local Similarity 48.0%; Pred. No. 7.3; Indels 0; Gaps 0;
Matches 98; Conservative 0; Mismatches 106;

QY 11 ATCTTTGACATCTACGTGTGTCACCGCTGACTACCTGCCCCCTAGGGGCTGAGCAGGATGCC 70
   |||||
Db 89 ACCITGTCGCATCGCGGACACGGTGTTCACCGCGCGCGGAGGCGGTGAATCGC 148

QY 71 ATCAGCTGCGGGAAGGCAGTATGTGAGTCTCTGATGACGCCACCCACTCCGCTGG 130
   |||||
Db 149 ACCCAGTCGCGCGTCAGCATGACATGAAGCGGCTCGAAGAGGACGTCCTGCAAGCGCTCG 208

QY 131 CTTGTCCGCACCAAGCCCAACCAAGTCCAGCCCTCAGCGCAGGCGCTGGGTGTCAACAGCC 190
   |||||
Db 209 CTGTTGAGCGCGGACGAGGACACCGCCTCACCGCCGAGGCCAGGTGCTCTCGGC 268

QY 191 TACCTGGACAGAGGCTCAAGCTG 214
   |||||
Db 269 TATGCCCGCGGATCCTCAAGCTG 292

RESULT 15
US-09-252-991A-13156/c
; Sequence 13156, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13156
; LENGTH: 1098
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13156

Query Match      8.9%; Score 34.4; DB 4; Length 1098;
Best Local Similarity 57.4%; Pred. No. 7.3;
Matches 62; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 273 ACGAATACAAGCAAGGCTGAGCTGTGTATCCAGGAGCTGCTGAGTCTGAGCAGGCT 332
   |||||
Db 911 ACGAAGCGCTTGTCCGGCCGATACTGCGATCCAGGCGTTGCGAGGCTCCGTCAGCTCC 852

QY 333 TCGTGGAGAGTGCAGTCTCTGAGAGCCACCACTGACCACTG 380
   |||||
Db 851 TCGAGCAGGTCGGCAGTCTCTCCAGGCAAGCGCCAGGTCCGCGAGG 804

Search completed: September 19, 2004, 22:12:45
Job time : 43.5752 secs
```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:07:04 ; Search time 1122.53 Seconds  
(without alignments)  
10242.013 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_16862\_17246

Perfect score: 385

Sequence: 1 ccgaggcgagatcttgaca.....acctgcagcactggagcgc 385

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rtd:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gss1:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110.6	28.7	512	12	B0059798
2	83	21.6	734	14	CF532602
C 3	74.8	19.4	528	12	BG792090
C 4	74.8	19.4	532	12	BG791767

5	43	11.2	507	9	AV388253
6	43	11.2	925	29	CNS0091P
7	42.6	11.1	965	13	BUS00169
8	42	10.9	428	9	AUI08162
9	40.8	10.6	881	13	BQ923183
10	40.4	10.5	628	12	BI527461
11	40	10.4	919	12	BI159893
12	39.8	10.3	846	29	CNS010RQ
13	39.6	10.3	935	29	CNS006XK
14	39.4	10.2	303	10	BF771180
15	39.4	10.2	347	28	AQ469228
16	39.4	10.2	354	10	BF893776
17	39.4	10.2	390	10	BF72502
18	39.4	10.2	390	12	BI060781
19	39.4	10.2	393	10	BF926934
20	39.4	10.2	444	10	AW484180
21	39.4	10.2	612	12	BG337262
22	39.4	10.2	710	12	BI823073
23	39.4	10.2	715	14	CA778105
24	39.4	10.2	825	13	BUS28667
25	39.4	10.2	849	12	BM018737
26	39.4	10.2	851	12	BG469835
27	39.4	10.2	873	13	BX357518
28	39.4	10.2	881	13	BQ80589
29	39.4	10.2	884	12	BI757129
30	39.4	10.2	903	13	BQ923314
31	39.4	10.2	910	12	BI757166
32	39.4	10.2	925	13	BUS28516
33	39.4	10.2	933	13	BQ896480
34	39.4	10.2	955	13	BUI75005
35	39.4	10.2	994	13	BX353043
36	39.4	10.2	995	13	BQ673197
37	39.4	10.2	1015	13	BQ058429
38	39.4	10.2	1201	13	BX417636
39	39.4	10.2	1214	13	BUI57066
40	39.2	10.2	1163	13	BX377454
41	39.2	10.2	1201	13	BX377844
42	39	10.1	816	14	CK139732
43	39	10.1	1201	29	CNS016BR
44	38.8	10.1	273	10	AW789770
45	38.8	10.1	401	14	HI9853

ALIGNMENTS

RESULT 1  
BJ059798  
LOCUS BJ059798 512 bp mRNA linear EST 29-SEP-2003  
DEFINITION BJ059798 NIBB Mochii normalized Xenopus tailbud library Xenopus laevis cDNA clone XL063014 5', mRNA sequence.  
ACCESSION BJ059798  
VERSION BJ059798.1 GI:17493138  
KEYWORDS EST.  
SOURCE Xenopus laevis (African clawed frog)  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.  
REFERENCE 1 (bases 1 to 512)  
AUTHORS Kitayama A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and Kohara, Y.  
TITLE Expressed genes in X. laevis embryo  
JOURNAL Unpublished (2001)  
COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp  
The information of this clone is available through the following URL.

```

http://xenopus.nibb.ac.jp.
Location/Qualifiers
1. 512
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL063014"
/tissue_type="whole embryo"
/dev_stage="stage 25"
/clone_lib="NIBB Mochii normalized Xenopus tailbud
library"

ORIGIN
Query Match 28.7%; Score 110.6; DB 12; Length 512;
Best Local Similarity 71.2%; Pred. No. 3.2e-14;
Matches 146; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 179 GTGTCCAGAGCTACCTGAGAGAGGCTCAAGCTGTCACCTGAGTGGGGGGCGCTGAG 238
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28 GGGTCTACCGCTTACTTGGATAAGAGACTGAAGCTATCATCTGAGTGGGCTGCAGGAGAA 87
QY 239 GCCCTGAGTTCCCTGGGGAGGCTGTGTCGAAGACGAATACAAGGCAAGGCTGAGCTCT 298
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
88 GCACAGAACTTCTGGAGTGTGTATCAGAAGAGAGGTACAGAAGAGAGCTCAGATT 147
QY 299 GTGATCCAGGAGCTGCTGAGTTTCAGAGAGCCCTTCGAGGAGCTGCAGTTCCTCCAG 358
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
148 CTGATCCAGGAGCTGCTAAACACAGAGAGGAGTACCTGAAGGAGCTACAGTCTTTCAG 207
QY 359 AGCCACCACTCCGACCTCGAGC 383
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
208 AACCATCACTACATCACTGAGC 232

RESULT 2
LOCUS CF532602 734 bp mRNA linear EST 12-SEP-2003
DEFINITION UI-M-GH0-cgw-a-11-0-UI.r1 NIH_BMAP_GH0 Mus musculus cDNA clone
IMAGE:30357226 5', mRNA sequence.
ACCESSION CF532602
VERSION CF532602.1 GI:34584570
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 734)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1. 734
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30357226"
/tissue_type="whole brain"
/dev_stage="1, 5, and 15 days newborn"
/lab_hosts="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_GH0"
/note="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I;

FEATURES
source
1. 528
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="UTSW H24F12"
/sex="Pooled"
/tissue_type="Cardiac muscle"
/lab_host="DH5a"
/dev_stage="2 months"

FEATURES
source
1. 528
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="UTSW H24F12"
/sex="Pooled"
/tissue_type="Cardiac muscle"
/lab_host="DH5a"
/dev_stage="2 months"

Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CGAAGTGAAT. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

ORIGIN
Query Match 21.6%; Score 83; DB 14; Length 734;
Best Local Similarity 86.0%; Pred. No. 4.9e-08;
Matches 92; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 35 GCTGACTACTCTCCCTAGGGCTGAGGAGGATGCATCAGCTCGGGAGAGCCAGTAT 94
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
628 GCTGACTATCTGCCACTGGAGCTGAGCAGGATGCCATCATTCGAGAGAGGCCAGTAT 687
QY 95 GTGAGAGTCTCTGATGTCAGCCACCACCTGCGCTGGCTTGTCCGAC 141
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
688 GTGAGAGTCTCTGATGTCAGCCACCACCTGCGCTGGCTTGTCCGAC 734

RESULT 3
LOCUS BG792090/c 528 bp mRNA linear EST 30-MAY-2001
DEFINITION UTSW H24F12 UTSW Adult Mouse Cardiac Muscle Library Mus musculus
cDNA clone UTSW_H24F12, mRNA sequence.
ACCESSION BG792090
VERSION BG792090.1 GI:14127660
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 528)
Gallardo,T.D., Schageman,J.J., Pertsemlidis,A., Garner,H.R.,
Williams,R.S. and Shchet,R.V.
UT Southwestern Medical Center, Adult Mouse Cardiac cDNA Library
Unpublished (2001)
Contact: Schageman JJ
Shohet/Garner Labs
University of Texas Southwestern Medical Center
6000 Harry Hines Blvd., NA2.226, Dallas, TX 75390, USA
Tel: 214 648 1674
Email: Jeff.Schageman@UTSouthwestern.edu
cDNA library constructed by UTSW as a component of the program for
Genomic Applications (PGA) and the Reynolds Heart Disease
Prevention Grants for use in cDNA microarray experiments. Sequence
Quality: Sequence ends were trimmed based on percentage of ambigu-
us base calls or 'N's in windowed segments. Sequencing: First-pass
sequencing; ABI prism 377 sequencer and analysis software.
Seq primer: M13/pUC Reverse.
Location/Qualifiers
1. 528
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="UTSW H24F12"
/sex="Pooled"
/tissue_type="Cardiac muscle"
/lab_host="DH5a"
/dev_stage="2 months"

FEATURES
source
1. 528
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="UTSW H24F12"
/sex="Pooled"
/tissue_type="Cardiac muscle"
/lab_host="DH5a"
/dev_stage="2 months"

note="Vector: pAMP10 (Gibco); Cloned unidirectionally.
Primer: Oligo dT. RNA isolation: cytoplasmic RNA preps
(Mannatis); Cloning Technique: CUA Cloning (Clontech),

```

<b>ORIGIN</b>					
Query Match	19.4%	Score 74.8;	DB 12;	Length 532;	
Best Local Similarity	77.4%;	Pred No. 2.8e-06;			
Matches 113:	Conservative				
	0: Mismatches	31:	Indels	2:	Gaps
					2:

## ORIGIN

240	CCCTTGAGTTCCCTGGGAGGCTGTCTGAAG-ACGNATCAAGGCAAGCTGAGCTCT	298
477	CCCCGAGTTCCTTGGTAAGGCTGTCTGAGGNATGAGTATAGACAGAGGCTGAGCTCT	418
299	GTGATCCAGGAGCTGCTGAGTCTTGAGCAGGCTTCGTGGAGGAGCTGCAGTTCCTGCAG	358
417	GTCAATCCAGGAGTTGCTGAGTTCAGAGCAGGCTTTGTGGTGANCNGCAGTCTCTTGAG	358
359	AGCCA-CCACCTGCAGCACCCTGGAGC	383
357	ANCCACCCACATGAAGACACNTGGAGC	332

## RESULT 5

AV388253	LOCUS	507 bp	mRNA	linear	EST 29-SEP-2000
DEFINITION	AV388253 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii				
CDNA clone	CM024c03_r, mRNA sequence.				
ACCESSION	AV388253				
VERSION	AV388253.1				
GI:	5542469				
KEYWORDS	EST.				
SOURCE	Chlamydomonas reinhardtii				
ORGANISM	Chlamydomonas reinhardtii				
	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;				
	Chlamydomonadaceae; Chlamydomonas.				

## FEATURES

```

source
1..507
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="CM024C03_r"
/dev_stage="photoautotrophic growth"
/clone_1i="Chlamydomonas reinhardtii C9"
/not_vector="false"
/site_1="EcoRI"
/site_2="XbaI"

```

**ORIGIN**

Query Match	11.2k;	Score 43;	DB 9;	Length 507;
Best Local Similarity	54.0k;	Pred. No. 31;		
Matches	88;	Conservative	0;	Mismatches 75; Indels 0; Gaps 0;
122	CTGCGCTGGCTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCSCAGGGGTGGTG	181		
37	CGGCGTGGCTTCTCGGCCAGCGCGGTGGGCCCGCGGCTTGCCTGCTC	96		
182	TCACCAAGCTACCTGGAGCAGGAGGCTCAAGGTGTCACTGAGTGGGGGGCGCTGAGGCC	241		
97	GTGCAGCGCTGGTGTGCGGGCTTCCACGCGGAGAGCCCTAGTGGGCTCGGTGACC	156		
242	CCTGAGTTCCTCGGGAGGCTGTCTCTGAAGACGAATACAAGG	284		
157	CCTGACATCAAGGCCAGGCCGCTTTCGACAGGAGTTCCAGG	199		



REFERENCE  
1 (bases 1 to 428)  
AUTHORS  
Sasaki,T. and Yamamoto,K.  
TITLE  
Rice cDNA from callus (2000)  
JOURNAL  
Unpublished (2000)  
COMMENT  
Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki  
305-8602, Japan  
Tel: 81-298-38-7441  
Fax: 81-298-38-7468  
Email: tsasaki@agr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/  
PROJECT = "RGP".

FEATURES  
source  
1..428  
Location/Qualifiers  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6426133"  
/clone="C50433"  
/clone\_lib="Rice callus"  
/note="Vector: pBluescript II SK+; Site 1: SalI; Site 2:  
NotI; cDNA prepared from rice callus mRNAs by using  
oligo(dT) as a primer and ligating to the SalI-NotI site  
of pBluescript II SK+ phagemid."

ORIGIN  
Query Match 10.9%; Score 42; DB 9; Length 428;  
Best Local Similarity 47.1%; Pred. No. 47;  
Matches 129; Conservative 0; Mismatches 145; Indels 0; Gaps 0;  
QY 111 CAGCCACCCACTCGCTGGTGTTCGACCAAGCCCAAGTCCAGCCCTCACGGC 170  
Db 53 CACCTCGCAGGTCCGAGCCTCTACCCACACCAAGCAGCAGCGTCCGTGC 112  
QY 171 AGGCTGGTGTACACGCTTACCTGGACAGAGGCTCAAGTCTACCTGAGTGGGG 230  
Db 113 TGGCGAGGTATCAGCACGTGAAGAGCTGAAGCGGCAGCAGCGGATCCGGCGG 172  
QY 231 CCGCTGAGGCCCTCGAGTTCCTGGGAGGCTGTCTGAGAGCAATACAAGCAAGGC 290  
Db 173 CAGCGCGCGGGGATTACACCGACAGCAGGAGCAGCAGATCGGTGTGGGGC 232  
QY 291 TGAGCTGTGATCAGAGAGTCTGAGTCTGAGAGGCTTCTGTGGAGAGCTGAGT 350  
Db 233 GCGGTTCGGCGCGCGCAGAGTGTCTGCGCAGCGAGCGGAGCAGCTGCGGTGACG 292  
QY 351 TCCTGCAGAGCCACACCTGCGACGACCTGGAGCG 384  
Db 293 CGCGGTGAGCGGAGGAGTCTGTGTGCG 326

RESULT 9  
BQ923183  
LOCUS  
DEFINITION  
AGENCOURT\_8803577 NIH\_MGC\_101 Homo sapiens cDNA clone IMAGE:6426133  
5' mRNA sequence.  
ACCESSION  
BQ923183  
VERSION  
BQ923183.1 GI:22339214  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 881)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM2609 row: a column: 14  
High quality sequence stop: 645.

FEATURES  
source  
1..881  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6426133"  
/tissue\_type="epidermoid carcinoma, cell line"  
/lab\_host="PH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_101"  
/note="Organ: lung; Vector: pOTB7; Site 1: EcoRI; Site 2:  
XhoI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

ORIGIN  
Query Match 10.6%; Score 40.8; DB 13; Length 881;  
Best Local Similarity 51.7%; Pred. No. 1.2e+02;  
Matches 93; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
QY 18 ACATCTACGTGTCACCGCTGACTCTCCCTAGGGCGCTGAGCAGGATGCATCACGC 77  
Db 576 AGCCTTCTGTGACGCGGTAAGTGCAGCTCTTTGCCCTCAAGGTGCTCAACATGA 635  
QY 78 TGGCGGAAGCCAGTATGTGGAGGTCTCGATGCAGCCACCCACTCGCTGTGTCC 137  
Db 636 TGCCCGAGGAGAGCTGGTTGAGGCCCTGCTGCAGCCACCGAGAAACAGAGAAAGCCC 695  
QY 138 GCACCAAGCCCAAGTCCAGCCCTCTACGCGAGGGCTGGGTGTCCAGGCTACCTGG 197  
Db 696 TGGAGAAGTGTCTCCCGGCTCTCTTGAGGCTGCTGAGCTGGTGCGACCTCCCTGG 755

RESULT 10  
B1527461  
LOCUS  
DEFINITION  
1024081609.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II  
Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION  
B1527461  
VERSION  
B1527461.1 GI:15368035  
KEYWORDS  
EST.  
SOURCE  
Chlamydomonas reinhardtii  
ORGANISM  
Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadaceae; Chlamydomonas.  
REFERENCE  
1 (bases 1 to 628)  
AUTHORS  
Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C.,  
Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.  
TITLE  
Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants. Project: 1024b  
JOURNAL  
Unpublished (2001)  
COMMENT  
Contact: Charles Hauser  
DCME Box 91000  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.

FEATURES  
source  
1..628  
Location/Qualifiers  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"





94	QY	TGTGGAGGCTCTGGATGTGACGCCACCACCGCTGCTGTGCGNCACGAGCCACCAA	153
658	Db	VSSSSSSAASASSASVSSSSSVAAVASASSSSSVSSSSAAVAAVVVSASSAC	717
154	QY	GTCACGCCCTCACGCGAGGCGGTGTCCACAGCCTACCTCGACGAGGAGGCTCAAGCT	213
718	Db	SASVSSVSGCAASSASASTSTSSAASASSSSSSGSSAGSSSSAASASSSSAAASA	777
214	QY	GTCACCTGATGGGGGGCGCTGAGGCCCTCGAGTTCTCTGGGAGGCGCTGTCTGAAGA	273
778	Db	ASSVSAVAVSASSSSASASSSSSSSSSSSSSSSSSSSSSASSASAAAASAVASSSSSGSAGSASS	837
274	QY	CGAATACAA	282
838	Db	SAGAAAATA	846

RESULT 13	CNS006XK/c
LOCUS	
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCSSION	AL066051
VERSION	AL066051.1 GI:4945019
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 935)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequences : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osagawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .

```

FEATURES
Source
Location/Qualifiers
1 935
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR14N09"
/clone_lib="RPCI-98"
/note="end : 177"

```

```

ORIGIN
Query Match 10.3%; Score 39.6; DB 29; Length 935;
Best Local Similarity 25.6%; Pred. No. 2.3e+02;
Matches 66; Conservative 70; Mismatches 112; Indels 0; Gaps 0;
/notes=end : 17.

```

[illegible]

148	QY	CACCAAGTCAGCCCTCAGCGAGGGCTGGGTGTACACAGCCTACTCTGACGAGGAGGCT	207
636	Db	CGCCSCGCGCGSCGCCGCCGCCCGCGAGCSCAGSCSCMGAVAGSGASRSVVGSGSS	577
208	QY	CAAGCTGTCACTGTGTGGGGGGCGCTGAGGCCCTGAGTTCCCTGGGAGAGCTGTGTC	267
576	Db	SGSASGCGCGCMGRAGSGKMGAGSGSGRCGACSGSGSBSKRGKGGAGCASCSSA	517
268	QY	TGAAGACG	275
516	Db	YKSGSCG	509

RESULT	14
BF771180	
LOCUS	BF771180 linear EST 12-JAN-2001
DEFINITION	IIL5-IT0027-271100-285-h08 IT0027 Homo sapiens cDNA, mRNA sequence.
ACCESSION	BF771180
VERSION	BF771180.1 GI:12119080
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 303)
REFERENCE	Dias Neto,E., Garcia Correa,R., Varjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Bretani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
AUTHORS	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
TITLE	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
PUBMED	10377800
COMMENT	Contact: Simpson A.J.G.

Contact: Rangelson A. de O. G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&ct=IL5-IT0027-271100-285-H08&t3=2000-11-27&ct4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 302.

```

FEATURES
source
    Location/Qualifiers
        1..303
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /dev_stage="Adult"
            /clone_lib="IT0027"
            /note="Organ: epid tumor; Vector: puc18; Site 1: SmaI;
            Site2: SmaI; A mini-library was made by cloning products
            derived from ORESTES PCR (U.S. Letters Patent application
            No. 196,716 - Ludwig Institute for Cancer Research)
            profiles into the pUC 18 vector. Reverse transcription of
            tissue mRNA and cDNA amplification were performed under
            low stringency conditions."

```

ORIGIN

Query Match 10.2%; Score 39.4; DB 10; Length 303;  
Best Local Similarity 51.4%; Pred. No. 1.5e+02;  
Matches 9; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Qy 18 ACATCTACGTGGTCAACGGTACTCTGCGCCCTAGGGGCTGACGAGGATGCCATCACGC 77

Job time : 1127.53 secs

```

Db      4  ACGCCTTCGTGAGCAGCGCTAAGTGCCAGCTCCTCTTTGCCCCCAAGGTGCTCAACATGA 63
QY      78  TCGCGGAAGGCGAGTATGTGGAGGTCTCTGGATGCGCCACCCACCTGGCTGGCTTGTCC 137
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      64  TCCCGGAGAGAGCTGTTGAGGCCCTGGCTGAGCCACCGAGAAGCAGAGAGAGGCC 123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      138  GCACCAAGCCACCAAGTCCAGCCCTCACGGCAGGGCTGGGTGTACACGACCTTACC 194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      124  TGGAGAAGTTGCTCCCGGCTCCTCTTGAGGGCTGCCTGAGCTGGTGGCACCCCTCCC 180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
AQ469228/c
LOCUS   AQ469228               347 bp    DNA        linear    GSS 23-APR-1999
DEFINITION   CITBI-EI-2601M16.TR CITBI-EI Homo sapiens genomic clone 2601M16,
              genomic survey sequence.
ACCESSION   AQ469228
VERSION     AQ469228.1  GI:4653118
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 347)
AUTHORS     Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
TITLE       Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
JOURNAL     Unpublished (1997)
COMMENT     Other GSSs: CITBI-EI-2601M16.TF
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.

FEATURES             Location/Qualifiers
     source            1..347
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2601M16"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-EI"
                     /note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     Caltech Human BAC Library D"

ORIGIN
Query Match      10.2%; Score 39.4; DB 28; Length 347;
Best Local Similarity 51.4%; Pred.No.1.6e+02;
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY      18  ACATCTACGTGGTCACCGCTGACTACCTGCGCCCTAGGGGCTGAGCAGGATGCCATCAGCG 77
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      194  ACGCCTTCGTGAGCAGCGGCTAAGTGCCAGCTCTCTTTGCCCTCAAGGTGCTCAACATGA 135
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      78  TCGCGGAAGGCCAGTATGTGGAGTCTCTGGATGCGCCACCCACCTGCGCTGGCTTGTCC 137
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      134  TCCCGGAGGAGAGCTGTTGAGGCCCTGGCTGCGCCACCGAGAAGCAGAGAGAGGCC 75
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      138  GCACCAAGCCACCAAGTCCAGCCCTCACGGCAGGGCTGGGTGTACACGACCTTACC 194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      74  TGGAGAAGTTGCTCCCGGCTCCTCTTTGAGGGCTGCCTGAGCTGGTGGCACCCCTCCC 18
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:07:04 ; Search time 3119.75 Seconds  
(without alignments)  
10242.013 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_1\_1070

Perfect score: 1070

Sequence: 1 tgctaccagagccacac.....tagtgcgcgagccgcggtt 1070

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: em\_estba:\*
- 2: em\_esthm:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hic:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gss1:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	260	24.3	911	BU180714	BU180714 AGENCOURT
C 2	249.8	23.3	728	AG034556	AG034556 Pan trogl
C 3	188	17.6	945	CNS04AD2	AL281711 Tetraodon
C 4	169.4	15.8	2650	AK031074	AK031074 Mus muscu

5	168.8	15.8	904	13	BU144883
6	167	15.6	558	11	AK029863
7	159.6	14.9	2710	11	AK034852
8	147.8	13.8	496	14	CA368743
9	144.8	13.5	752	29	CNS01YES
C 10	127.4	11.9	539	9	AL927364
11	124.4	11.6	1136	12	BM799726
12	122.2	11.4	929	13	BO679064
13	122.2	11.4	1017	13	BO679833
14	119	11.1	910	12	BA908078
15	112.8	10.5	407	10	AM170791
C 16	112.2	10.5	437	29	CE070659
17	111.2	10.4	718	13	BO770422
18	108.8	10.2	780	29	CNS04142
19	107.8	10.1	534	14	CD555386
20	106.8	10.0	519	12	BI789591
21	106.6	10.0	518	10	BE648756
22	106.2	9.9	756	14	CB249772
23	106.2	9.9	4130	11	AK048144
24	99.6	9.3	947	29	AG069825
25	98.4	9.2	848	29	AG032377
C 26	98	9.2	1165	29	AG030649
C 27	98	9.2	1542	29	AG032943
C 28	97.6	9.1	1598	29	AG030579
C 29	96.8	9.0	1625	29	AG043477
30	96.6	9.0	1516	12	BG809984
31	94.8	8.9	1026	28	BZ569417
32	94.4	8.8	932	29	CNS0072Q
C 33	94.4	8.8	1610	28	BZ569386
C 34	94.2	8.8	1452	29	AG032979
35	94	8.8	935	29	CNS006XK
36	94	8.8	982	13	BX415111
C 37	92	8.6	982	13	BX415111
38	92	8.6	1124	29	AG041123
39	91.8	8.6	1085	29	AG152796
40	91.2	8.5	454	10	AW410057
41	91	8.5	1189	29	AG030608
42	91	8.5	1341	29	AG030611
C 43	90.2	8.4	863	12	BG809696
44	90.2	8.4	1070	28	BZ557810
45	90.2	8.4	1116	12	BG810038

ALIGNMENTS

RESULT 1  
BU180714/c  
LOCUS BU180714 911 bp mRNA linear EST 04-SRP-2002  
DEFINITION AGENCOURT\_7968637 NIH\_MGC\_68 Homo sapiens cDNA clone IMAGE:6011039  
5', mRNA sequence.  
ACCESSION BU180714  
VERSION BU180714.1 GI:22694698  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 911)  
AUTHORS NIH-MGC http://imgc.ncbi.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DP/Gazdar  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM13200 row: a column: 24  
High quality sequence stop: 409.

BU144883 603229681  
AK029863 Mus muscu  
AK034852 Mus muscu  
CA368743 645003 NC  
AL172886 Tetraodon  
AL927364 AL927364  
BM799726 AGENCOURT  
BO679064 AGENCOURT  
BO679833 AGENCOURT  
BG908078 602520064  
AG170791 xj25b01.x  
CE070659 tigr-ges-  
BO770422 UI-M-F10-  
AL291788 Tetraodon  
CD555386 B0391F02-  
BI789591 ic96f10.y  
BE648756 BB648756  
CB249772 UI-M-EXO-  
AK048144 Mus muscu  
AG069825 Pan trogl  
AG032377 Pan trogl  
AG030649 Pan trogl  
AG032943 Pan trogl  
AG030579 Pan trogl  
AG043477 Pan trogl  
BG809984 mgct002xd  
AL066742 Drosophil  
BZ569417 pac62-164  
BZ569386 pac62-164  
AG032979 Pan trogl  
AL066051 Drosophil  
BX415111 BX415111  
BX415111 BX415111  
AG041123 Pan trogl  
AG152796 Pan trogl  
AW410057 fh03h11.x  
AG030608 Pan trogl  
AG030611 Pan trogl  
BG809696 mgct001xf  
BZ557810 pac61-60  
BG810038 mgct002xi



Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodontidae; Tetraodon.

1  
Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,  
Bernot, A., Frazzetta, C., Wincker, P., Brothier, P., Quetier, F.,  
Saurin, W., and Weissenbach, J.  
Estimate of human gene number provided by genome-wide analysis  
using Tetraodon nigroviridis DNA sequence  
Nat. Genet. 25 (2), 235-238 (2000)  
20296633  
10835645

2  
Roest Crolius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,  
Frazzetta, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,  
Saurin, W., Bernot, A., and Weissenbach, J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish tetraodon nigroviridis  
Genome Res. 10 (7), 939-949 (2000)  
20359837  
10899143

3 (bases 1 to 945)  
Direct Submission  
Genoscope.  
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/tetraodon.  
Location/Qualifiers  
1..945  
/organism="Tetraodon nigroviridis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:99883"  
/clone="095C14"  
/clone\_lib="G"  
/note="Genoscope sequence ID : C0B095B0879P1-end :  
PUC-ori"

source  
Query Match 17.6%; Score 188; DB 29; Length 945;  
Best Local Similarity 56.7%; Pred No. 6.6e-16;  
Matches 388; Conservative 1; Mismatches 286; Indels 9; Gaps 2;

QY 78 CAGGCACAGTTAGCGGGCGCCCGCTTTCTACCCGGCCCAAGGCTTCGTGTGTCG 137  
DB CACACCTATTGGCGGAGTCCCMAGGTTCTAACCGCGCCGAAGGCTTTTCGTGTGT 834

QY 138 GTGGGAGGAGCGCCACCTCAGCTGCAGATCTGGGTATCCACGCCACAGTGAGC 197  
DB GTGGGCAAGATGGACTCTCAGCTGCACTGTGTGGAAATCGACCCCGCTGATAACC 774

QY 198 TGGGAGAGGAGCAGCAGCGCGGTGACCGCGCGCGCTTCCTGCTGCGCCAGGACGGC 257  
DB TGGGAAGGAGCAAGCTGAAGCTGACCTCTGGGGAGCATCAAGACCGTGGAGATGA 714

QY 258 GACCTTACCGCTCAGTATCTGAGACTGTGGCGTGGCGGACAGTGGCGCAATACGTGTC 317  
DB GATGTGTACCGCTGACCATCTACGAGCTGACCCCTGGAAGACAGCGCTCAGTATGTGC 654

QY 318 CGCGCGCGCAATGTCATAGGCGAGCGCTTCGCTCCGCTGGGCTTCGAGGTGGACGGGAG 377  
DB AGAGCCAGACACAGTTGGTGAAGCTAGCGCGCGCTGACCTCAAGTGGCGCTCCG 594

QY 378 CGCGGTGCGCGAGCAGGCGCCGCACTTCTGCTGCGGCGCAAGTCCATCCCGTGC 437  
DB GCAGAGATGCCCAAGAGGCGCCCTGTCTGCTGCTCAAGCTGCGACCGCGCGTGTAGC 534

QY 438 GAGGGCTCAGAGGCCACCTTCGCTGCGCGGTGGTGGTCCCGAGCGCGCAGTGAGC 497  
DB CTTGGAGAGAGCGTGTCTGCTGCTGCGGAGTCCAGCTCACCCTCCGAGCCAACTTGAA 474

QY 498 TGGTCCAAAGGACGCGGCGCGCTTGGGTGAGCCGACGCGCGCGCTGCGCGTGGAGGAG 557  
DB TGGGAGAGGACGCGGACGCTTACCTTAGGGAG-AGCAATCGCGTCAAGATTGTTC 420

QY 558 CTGGCGGAGCAAGTGGCTGCGCATTCGGGCGCGCGCGCGCGCGCGCGCGCACTTAC 617  
DB GACGCGGACAGCAGCATTCGAAATCCAAAGGGTGGTCACTTGGCGCGCGCACTTAC 360

QY 618 GAGGTCCGCGCGCGAGAACCGCTGGGCGGTGCGCGCGCGCGCGCGCGCTAGTGTGGAC 677  
DB ACCTGCGAGGCGCGAGAACCGCTGGGCGGTGCGCGCGCTGCGCGCGCTGCGCGTGGAC 300

QY 678 TCGGAGCGCGCGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCTGGCGCAC 737  
DB GCCAGATTCACAGCGCTGATAAACCAGATAAAGCA---CCTCCCTGCTGTCTCAC 243

QY 738 CTGCGAGCGCGCGCGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCTAGTGTGGAC 761  
DB CTTCAAAGCGCGCAAGAGGAATG 219

RESULT 4  
AK031074  
LOCUS  
DEFINITION  
AK031074 2650 bp mRNA linear HTC 18-SBP-2003  
Mus musculus 13 days embryo forelimb cDNA, RIKEN full-length  
enriched library, clone:5930402G23 product:hypothetical  
Immunoglobulin and major histocompatibility complex domain  
containing protein, full insert sequence.  
AK031074  
ACCESSION  
VERSION AK031074.1 GI:26326994  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636

2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20493374  
11042159

3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
Yoneda, F., Ishikawa, F., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

4  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

5  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs



KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 904)  
AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
TITLE A Comprehensive Collection of Chicken cDNAs  
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
MEDLINE 22335534  
PUBMED 12445392  
COMMENT Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.  
FEATURES  
source Location/Qualifiers  
1..904  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="Compton Line 151"  
/db\_xref="taxon:9031"  
/clone="CHST223421"  
/sex="Female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="CSEQC126"  
/note="Organ: heart; Vector: pBluescript II KS(+); Site 1:  
EcoRI; Site 2: NotI; Modification of pBluescript II KS(+)  
[Stratagene] vector to accommodate cDNA produced with the  
T-trimmed protocol (Construction of uni-directionally  
cloned cDNA libraries from messenger RNA for improved 3'  
end DNA sequencing by Glenn Fu, et al. U.S. Patent #  
6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI.  
Ligate in double stranded adaptor containing BspI and  
BamHI sites [5'ggccggtgcagcccgatccgaaaaag]  
[5'aattcttttttcggatccggggtgcagc]"

## ORIGIN

Query Match 15.8%; Score 168.8; DB 13; Length 904;  
Best Local Similarity 72.7%; Pred. No. 2.5e-13;  
Matches 218; Conservative 0; Mismatches 82; Indels 0; Gaps 0;  
QY 764 CGCCGAGGCGCCCGCCCTCACGCCAGCCGCGACCGCGACCTGCGACGGTACTGA 823  
Db 29 CACCAAGAGGGCTCGCAATGCCATGTTGGACGCTGACGCGCACCTGCGACGGA 88  
QY 824 AGGCAAGCACGCGCGCCCTCAGTGTGCTACGTGACCGGGGAGCCCAAGCCGAGCGGTGTG 883  
Db 89 AGGGAAGCACGCGCAAGCTGACTGCTAGTGTGCTGGGAGCTGAAGCCAGATCGTGTG 148  
QY 884 GAAGAGAGCGCGCGAGCTGTGTGACCGAGGGCGCGCCAGCGTGTGTACGAGACGCGCA 943  
Db 149 GAAGAAGGACCAACGAGGTCTATTGTGAAGGGCGGAGGCGATGCTATCTATGAGATGATCA 208  
QY 944 GGAGAACTTCGTGCTCAAGATCTCTTCTGCAAGCAGTCGACCGCGGCTCTACACCTG 1003  
Db 209 GGAGAACTTTCGTGCTGAAGATCTCTTCTGCAAGCAGTGAACCGGCTGTACACCTG 268  
QY 1004 CACGGCGTCCAACTCGTGGGCCAGACCTACAGCTCTGTGCTGTGTGTGCGCGGACCC 1063  
Db 269 CACGGCGTCCAACTCGGCGGGCAGACCTACAGCTCTGTGCTGTGTGTGCTCACTGTCAAAGAAC 328

## RESULT 6

AK029863  
LOCUS 5558 bp mRNA linear HTC 18-SEP-2003  
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched

library, clone:4931417C03 product:hypothetical Immunoglobulin and  
major histocompatibility complex domain containing protein, full  
insert sequence.  
AX029863 1 GI:26081564  
VERSION AK029863 1 GI:26081564  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
REFERENCE 3  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kusunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Hara, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Hashida, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohata, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861  
REFERENCE 4  
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)  
PUBMED 11076861  
REFERENCE 5  
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
PUBMED 120530913  
REFERENCE 6  
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Hayashida, K., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hori, F., Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, K., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
COMMENT cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues. Please visit our web site for further details. URL: <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> URL: <a href="http://fantom.gsc.riken.go.jp/">http://fantom.gsc.riken.go.jp/</a>		Db		649		CCCCAACGCGCACGGCCACGCGCAGCGCGCGCGCTGTCTCCAGGTGCACCCAGCCCGCGC		708		
FEATURES		Qy		741	Db	CAGCGCGCGCGAGGCTATCGCGCGCGAGGCGCGCGCGCTCACCGCCAGCAGCCCGC		800		
Location/Qualifiers		Db		709		GAGAGCGCGCGCGAGGATCCCGATGAGAAACCCCAACCCGCTTTGGAGCGCGCTCAAGGC		768		
1. 5558		Qy		801		ACGCGC---ACCTGCACGGTGAAGGCAAGCAGCGCGCGCTCAGTGTCTAGCTGACC		857		
/organism="Mus musculus"		Db		769	Qy	GCGCCCAAGACCTTCTGGGTGAACGAGGCGCAAGCAGCGCGCTTCCGCTGTCTAGCTGATG		828		
/mol_type="mRNA"		Qy		858		GGCGAGCCCAAGCCCGAGAGCGGTGTGGAAGAAGAGCAGCGCGCTGTCACCGAGGCGCGG		917		
/strain="C57BL/6J"		Db		829		GGCAAACTGTAGCCCGAGATCGAATGGCACTTTGGAGGCGCGCGCTCTGTCTCCCGATCGC		888		
/db_xref="PANTOM DB:4931417C03"		Qy		918	Qy	CGCACGTGGTGTACAGAGCGCGCAGGAGAACTTCGTCTCAAGATCTCTTCTTGCAGAG		977		
/db_xref="MGI:2392423"		Db		889		CGCGCGCTCATGTACCGCGACCGCGCGCGCTTTGTACTTAAGGTGCTCTACTGCGCAG		948		
/db_xref="taxon:10090"		Qy		978		CAGTCGCGCGCGCGCTTACACCTGACGCGCGCTCCAACTCTGCGGCGCAGACCTACAGC		1037		
/clone="4931417C03"		Db		949	Qy	GCCAAGGACCGTGGGCTCTACGTTGTGCGCGCGCGCACTCGCGGCGCGCAGACCTAAGC		1008		
/sex="male"		Qy		1038		TCTGTGCTGGTCTAGTGTGCGCGAGGCC		1064		
/tissue_type="testis"		Db		1009		GCGGTCCAGCTGCACGTGAAAGAACCC		1035		
/clone_lib="RIKEN full-length enriched mouse cDNA library"		misc_feature		polyA_signal		5534. 5539		polyA_site		
/dev_stage="adult"		polyA_site		5558		5558		ORIGIN		
/dev_stage="adult"		Query Match		15.6%; Score 167; DB 11; Length 5558;		Best Local Similarity		50.5%; Pred. No. 6e-13;		
/note="hypothetical Immunoglobulin and major histocompatibility complex domain containing protein (InterPro IPR003006, evidence: InterPro) putative"		Matches		498; Conservative		0; Mismatches		470; Indels		
/note="putative"		Qy		97	Db	CGCCCGCTTCTCACCGCGCCAGGCTTCGTGGTGGTGGCAAGGAGCGCCACCC		156		
/note="putative"		Qy		49		CCCCGTGTTCTACGCTTCCCGCGCGCGTGGGTTGAAGTGGAGCGGAGCGCGAC		108		
/note="putative"		Qy		157		TCAGTCGCGATGTGGTAAATCCACGCGCACAGGTGAGTGGGAGAGGACCGACGAC		216		
		Db		109	Qy	TCAATGGTGTCTGGAGAGCGCGCGCCACTTCGTGTGGAGAAAGGCGGCGAGC		168		
		Qy		217		CGGTGACGCGCGCGCGCTTCGTCTGGCGCCAGGACGCGCACTTACCGCTCACTA		276		
		Db		169		AGTTGGTGGCTCCGAGCGCTAAGCTTTCCGGAGGACGCGCGCGCGCTGCTGC		228		
		Qy		277	Db	TCTGGAACCTGCGTGGGCGACAGTGGGCAATACGTGTGCGCGCGCGCAATGCCATAG		336		
		Qy		229		TGAGCGCGCGCTGCCACCGACCGCGGGTCTACGTGTGCGCGCGCGCAACGCGCGC		288		
		Db		337		CGGAGCGCTTC-----GCTGCGCTGGCGCTGCAGGTGGACGCGGAGGCGCGTGGCGCG		390		
		Qy		289	Db	GAGAGCGCTACGCGCGCGCGCTTACCGTGTGGAACCCCGCGCGCGCGCGAGC		348		
		Qy		391		AGCAGCGCGCGCACTTCTGCTGGCGCGCGCGCTCCATCCGCTGCGCGAGGGCTCAGAGG		450		
		Db		349		CGAGTCTCCGAGTGTCCGCTGCAACACCGCGCACCGGCGGAGGCGCGCGCAATTC		408		
		Qy		451	Db	CCACCTTCGCTGCGCGCTGGTGGCTCCCGAGCGCGCGAGTGTGTCGCAAGGACG		510		
		Qy		409		TGACGGGCGCGCGAGTCCCGAGTGGGTGCTGCGAGGGGAGGAGTGGTGTGCTGACGTGCCAGG		468		
		Qy		511		GCGCGCGCTGGGTGAGGCGCGCGCGCGCGCGCTGCGCTGGAGAGTGTGGCGAGGCA		570		
		Db		469	Qy	TGGGAGGCGCTTCCGAGGCGCGCGAGCTGTACTTGGGAGAGGATGGATGGCTTGGACGAG		528		
		Qy		571		GTGCGCTGCGCATTCGCGCGCGCGCGCGCGCGCGCTTACGAGTCCGCGCGCGCGCGCG		630		
		Db		529		TATGGGACAGCACCCACTTCAAGCTGGAGCGCGCGCGCGCGCGAGTACGAGGCGCGA		588		
		Qy		631	Db	AGAACCGCGTGGCGCT-----GCCAGCGCGCGCGCGCGCGCTAGTGTGAGCTCG		680		
		Qy		589		GCCTGACGCTTGGCATCTCGCGCGCACGCGTGCGCCGATTCGCGGGTGTACGTGTGTCA		648		
		Db		681		GACGCGCGGACACGCGCGCGCGCGCGCGCTCCACGCGCGCGCTCTCTGGCGCACCTG		740		





```

VERSION      CA368743.1  GI:24680400
KEYWORDS     EST.
SOURCE       Oncorhynchus mykiss (rainbow trout)
ORGANISM     Oncorhynchus mykiss
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Actinopterygii; Neopterygii; Teleostei; Euteleostei;
              Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
TITLE        1 (bases 1 to 496)
JOURNAL      Rextroad, C.E. and Keele, J.W.
COMMENT      Sequence analysis of a rainbow trout normalized cDNA library
              Unpublished (2002)
              Contact: Rextroad CE
              USDA, ARS, National Center for Cool and Cold Water Aquaculture
              11876 Lestown Road, Kearneysville, WV 25430, USA
              Tel: 304 724 8340 X2129
              Fax: 304 725 0351
              Email: crexroad@cccwa.ars.usda.gov
              Single pass sequencing. Bases called with phred v0.020425.c and
              trimmed with the aid of the trim_alt option. Vector identified by
              cross match v0.990329.
              Seq primer: AGCGATACCAATTTCACACAGGA.
              Location/Qualifiers
FEATURES     source
              1..496
                /organism="Oncorhynchus mykiss"
                /mol_type="mRNA"
                /db_xref="taxon:8022"
                /clone="1RT155J08 D E04"
                /tissue_type="pooled"
                /lab_host="DH10B"
                /clone_lib="NCCOWA 1RT"
                /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
                Library made from pooled tissue from brain, gill, liver,
                spleen, muscle, and kidney."
ORIGIN
Query Match      13.8%; Score 147.8; DB 14; Length 496;
Best Local Similarity 68.0%; Pred. No. 1.5e-10;
Matches 206; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY      69  GTCTATGATCAGCCACAGTTCAGCGGGCGCCCGCTTCTCACCAGCCGCGCCCAAGGCTTC 128
Db      179  GTATGACACAGACCTGTTTGGAGGGCGCCCGAGGTTTCTCAGCGGCCCAAGGCTTC 238

QY      129  GTGTGTGGTGGGCAAGGACGCCACCTCAGCTGCGCAGATCGTGGGTAAATCCACGCCA 188
Db      239  GCAGTGTGCGTGGGAAAGGATGCCACCTTGAGCTGCACCATCGTGGGCAACCCCAACCCCA 238

QY      189  CAGGTGAGCTGGGAGAGGACACGACCGGTGACGCGCGCGCGCTTCCGCTCTGGCC 248
Db      299  CTGATACCTGGGAAAGGAGAACTAGCTGACCTCTGGGGCCGCTTCAAGATGGTG 358

QY      249  CAGGACGCGCACCTCTACCGCTCCTACTATCTGACACCTGCGCTGGGCGCAGATGGGCA 308
Db      359  GATGATGTGACGTCTACCGCTCCTACCATCTACGATCTAAACCTTGGAGGACAGCGGCCAG 418

QY      309  TACGTGTCCCGCGCGCATGCGCATAGGCGAGCGCTTCCGCTGCGGCGCTGCGAGGTG 368
Db      419  TACATGTCCCGCGCGCAAAACAATGTTGGGGAACATATGATGCTGTAACCTTTCAGTA 478

QY      369  GAC 371
Db      479  GGC 481

RESULT 9
CNS01YE5    752 bp  DNA  linear  GSS 01-SEP-2000
LOCUS       Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
DEFINITION  218A06 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   ALI72886
VERSION     ALI72886.1  GI:7810943
KEYWORDS    GSS; genome survey sequence.

```

```

SOURCE       Tetraodon nigroviridis
ORGANISM     Tetraodon nigroviridis
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Actinopterygii; Neopterygii; Teleostei; Euteleostei;
              Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
              Tetraodontoidea; Tetraodontidae; Tetraodon.
TITLE        1
JOURNAL      Roost Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
              Bernot, A., Fzames, C., Wincker, P., Brottier, P., Quetier, F.,
              Saurin, W. and Weissenbach, J.
COMMENT      Estimate of human gene number provided by genome-wide analysis
              using Tetraodon nigroviridis DNA sequence
              Nat. Genet. 25 (2), 235-238 (2000)
              20296633
              MEDLINE
              PUBMED
              10835645
AUTHORS      Roost Crollius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,
              Fzames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
              Saurin, W., Bernot, A. and Weissenbach, J.
TITLE        Characterization and repeat analysis of the compact genome of the
              freshwater pufferfish Tetraodon nigroviridis
JOURNAL      Genome Res. 10 (7), 939-949 (2000)
MEDLINE      20359837
PUBMED       10899143
AUTHORS      3 (bases 1 to 752)
TITLE        Direct Submission
JOURNAL      Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail : secre@genoscope.cns.fr)
              Web : www.genoscope.cns.fr)
COMMENT      This sequence is a single read and was generated as part of a large
              scale clone-end sequencing project of the Tetraodon nigroviridis
              genome. For more information, please take a look at
              http://www.genoscope.cns.fr/Tetraodon.
FEATURES     source
              1..752
                /organism="Tetraodon nigroviridis"
                /mol_type="genomic DNA"
                /db_xref="taxon:99883"
                /clone="218A06"
                /clone_lib="G"
                /note="Genoscope sequence ID : COAG218BA039P1-end :
                PUC-ori"
ORIGIN
Query Match      13.5%; Score 144.8; DB 29; Length 752;
Best Local Similarity 60.9%; Pred. No. 4e-10;
Matches 284; Conservative 1; Mismatches 178; Indels 3; Gaps 3;

QY      71  CATGGATCAGCACAGTTTCAGCGGGCGCCCGCTTCTCACCAGCCGCGCCCAAGGCTTCGT 130
Db      232  CATGATCCACACCTATTTCGGGAGTCCCGAGTTTCTAACGC-GCCCGAAGCTTTTC 290

QY      131  GGTGTGCGTGGGCAAGGACGCCACCTCAGCTGCCAGATCGTGGGTAAATCCACGCCACA 190
Db      291  GGTGTGTTGGGCAAGATGCGACTCTTAGCTGCACTGTGTGGGAAAGTCCGACCCCGCT 350

QY      191  GGTGAGCTGGGAGAGGACGACCGGTGACGCGCGCGCGCGCTTCGCTCTGGGCCA 250
Db      351  GATAACCTGGGAAAGGGAAGCTGAAGCTACGCTCTGGGGGACGATTCAAGACGTTGGA 410

QY      251  GGACGGGACCTCTACCGCTCCTACTATCTCTGAGCTGGGCGCTGGGCGACAGTGGGCAATA 310
Db      411  GGATGGAGATGTGTACCCGCTGACCATCTACGAGTGAACCTGGAAGACAGCGGTCACTA 470

QY      311  COTGTGCGCGCGCGCAATGCGATAGGAGAGCTTCGCTCCGCTGGGCTTGCAGGTGGA 370
Db      471  TATGTACAGAGCAAGAACAAACGTTGTGTGAAGCGTAGCTCCGCTTACCTCTCAAGGTGGC 530

QY      371  CGCGGAGCGCGGTGCGCGCGGACGAGCGCGCGCTTCCTGCTGGCGGCCACGCTCCATCCG 430
Db      531  CTTGCGCGCAGATGCCCCAG-AGGCCCTCTCTTCTGCTCAAGCTGCCACCGCGCG 589

```

QY 431 CGTGGCGAGGGCTCAGAGGCCACCTTCGGTCCGGTGGGTGGCTCCCGAGGCGGC 490  
Db 590 TTTAGGCTGGAGGAGAGCTTCTTCGTCGTCGAGTCGACCTCACCCGGCCAC 649  
QY 491 AGTGAAGTGTCCAGAGCGCGCGCGCTGGTGGTGAGCCGACGCG 536  
Db 650 TTTGA-ATGGGAGAAGGACGACGCTACCTAGGGGAGAGCAATGCG 694

RESULT 10  
AL927364/c  
LOCUS AL927364 PJR-Z1-Z2 Danio rerio cDNA clone 188-F02-2, mRNA sequence.  
ACCESSION AL927364.1 GI:23197401  
VERSION AL927364.1  
KEYWORDS EST.  
SOURCE Danio rerio (zebrafish)  
ORGANISM Danio rerio

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 539)

AUTHORS Lo, J., Lee, S., Xu, M., Liu, F., Ruan, H., Eun, A., He, Y., Ma, W.,  
Wang, W., Wen, Z., and Peng, J.  
TITLE 15000 unique zebrafish EST clusters and their future use in  
microarray for profiling gene expression patterns during  
embryogenesis  
JOURNAL Genome Res. 13 (3), 455-456 (2003)  
MEDLINE 22505427  
PUBMED 12618376

COMMENT Contact: Peng J  
Lab of Functional Genomics  
Institute of Molecular and Cell Biology  
30 Medical Drive, Singapore, 117609, Singapore  
Email: pengj@imcb.a-star.edu.sg  
Clone requests: pengj@imcb.a-star.edu.sg.  
FEATURES  
source  
1..539  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/strain="local wildtype"  
/db\_xref="taxon:7955"  
/clone="188-F02-2"  
/tissue\_type="whole embryo or fish"  
/dev\_stage="mixed stages"  
/clone\_lib="PJR-Z1+Z2"

ORIGIN  
Query Match 11.9%; Score 127.4; DB 9; Length 539;  
Best Local Similarity 60.1%; Pred. No. 8.3e-08;  
Matches 212; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 17 ACATCTCGGCGCTGCCAGAGCCCCCATAGAGAGGTCCCGCCCGCATGGA 76  
Db 470 AAACATCTGCACAGTTAATATCATCTATATTTGAGGAGGAAAGAGCAATGGA 411

QY 77 TCAGCCACAGTTACGGGGGCGCCGCTTCTCACCGCGCCCAAGGCTTCGTGGTGC 136  
Db 410 CCAGATCTATTGGGAGAGTCTCTGTTTCTTCTGACCCCAAGGCTTCTTTGTG 351

QY 137 GGTGGGCAAGACGCCACCTTCAGTCCAGATCGTGGTAAATCCCAAGCCACAGGTGAG 196  
Db 350 TGTGGAGCGGACGCGCTCTCTTAGCTGCACCTATTGTGGAAACCCCTGTACTGTAGTCAC 291

QY 197 CTGGGAGAGACACGACGCGGTGACGCGCGCGCGCTTCGCTGCGCCAGGACGG 256  
Db 290 CTGGGAGAGAGAAATTCGCTATTCTGCGAGGGGGCGGATTCAAAATCTGTCGAGGATGG 231

QY 257 CGACCTCTACCGCTCTACTATCTCGGACCTCGGCGTGGCGACAGTGGGCAATACGTGTG 316  
Db 230 TGACATTTATCGATTACTATCTATGATCTAACGCTAGAGACAGTGGTCAGTACATGTG 171

QY 317 CCGCGCGCGCAATGCCATAGCGGAGGCGCTTCGCTCGCGCGCTCGCGAGGTGG 369

Db 170 TCGTGCCAAAGACACCGTTGGAGAGGCATATGCAGCTGTAAACCTGAAGGTGG 118  
RESULT 11  
BM799726  
LOCUS BM799726  
DEFINITION BM799726 1136 bp mRNA linear EST 05-MAR-2002  
5', mRNA sequence.  
ACCESSION BM799726.1 GI:19116549  
VERSION BM799726.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1136)

AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM12122 row: p column: 11  
High quality sequence stop: 635.  
FEATURES  
source  
1..1136  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5495674"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 67"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site: 1: Notr;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 1.75 kb. Library constructed by Life  
Technologies."

ORIGIN  
Query Match 11.6%; Score 124.4; DB 12; Length 1136;  
Best Local Similarity 59.0%; Pred. No. 2.4e-07;  
Matches 237; Conservative 0; Mismatches 156; Indels 9; Gaps 1;

QY 672 GTGGACTCGAGCGCGGACACGCGCCAGCGCGGACCTCCACGCGCGGTCTCTG 731  
Db 125 GTCTAGTGTGCCACCGCGCAACGCGCACGCGCGGCGGCGGTCTCTCCAG 184

QY 732 GCGCACCTTCAGCGCGGCGCGAGGCTATGCGCGCGCGGCGCGCCCGCCTC- 784  
Db 185 GTGCACCGCGCGCGAGAGCGCGCGCGGACCGCGCGCGCGCGCGGTGGTG 244

QY 785 --ACGCGCCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 842  
Db 245 GAGCGCGCTCAAGTGGCGCGCTTAAGACCTTCTGGGTGACAGGGGCAAGCACCGCAAGTTC 304

QY 843 AGCTGTACTGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 902  
Db 305 CGCTGTCTAGTGTGGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 364

QY 903 GTGACCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCAAG 962  
Db 365 CTGCTCCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCAAG 424

QY 963 ATCTCTTCTTGAAGCAGTCGGACCGCGCGCGCTTACCTGACCGCGCGCGCTCGTG 1022  
Db 425 GTGCTTTACTGCAGCGCAAGGATGTGGGCTTCTAGCTGTGCGCGCGCGCACTCGGCG 484



QY 950 CTTCGTGCTCAAGATCTCTTCTCAAGCAGTCGAGCGCGGCTCTACACCTGCAGCGC 1009  
|||||  
Db 249 CTTCGTGCTCAAGGTGCTCTACTCCAGGCCAAGATCGTGGGCTCTACGTCGTGGCGCG 308  
|||||  
QY 1010 GTCCAACTCTGTGGCCAGACCTACAGCTCTGTGCTGTGTAGTGTGGCGAGGCC 1064  
|||||  
Db 309 GCGCAACTCGCGCGCCAGACGCTCAGTGGCGCTGAGCTGCACGTGAAGAGGCC 363  
|||||

RESULT 14  
BG490878  
LOCUS  
DEFINITION BG490878 910 bp mRNA linear EST 27-MAR-2001  
602520064P1 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:4638590 5',  
mRNA sequence.

ACCESSION  
VERSION BG490878  
KEYWORDS  
SOURCE EST.  
ORGANISM Homo sapiens (human)

REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 910)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov

Tissue Procurement: DCTD/PTP/Gazdar  
CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>  
Plate: LLM1400 row: p column: 15

High quality sequence stop: 802.

FEATURES  
Location/Qualifiers

1..910  
/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4638590"

/tissue\_type="large cell carcinoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_18"

/note="Organ: lung; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGGAG(G). Library constructed by Ling Hong in the  
Laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

ORIGIN

Query Match 11.1%; Score 119; DB 12; Length 910;  
Best Local Similarity 64.2%; Pred. No. 1.2e-06;  
Matches 179; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 786 CCGCCACGACCGGACCGGACCTGACGGTGAAGCAGCAGCGCGGCTCAGC 845  
|||||

Db 20 CCGCTCAAGTGGCGCGCTTAAGACCTTCTGGGTGAACGAGGCGAAGTTCGCG 79  
|||||

QY 846 TGCTACGTGACCGGCGAGCCCAAGCCGAGACGGTGTGGAAGAGAGCGCAGCTGGT 905  
|||||

Db 80 TGCTACGTGATGGGCAAGCCCGAGCCCGAGATCGATGCGACTGGAGGCGCGCGCTG 139  
|||||

QY 906 ACCGAGGCGCGCGCCACGTGGTGTACGAGGACGCGCAGGAGAACTTCGTGCTCAAGATC 965  
|||||

Db 140 CTCGCGGACCGCGCGCGCTCATGTATACCGGACCGGCGGCTTCGTGCTCAAGGTG 199  
|||||

QY 966 CTCCTTCGACAGCTCGGACCGCGGCTCTACACCTGCACGCGGTCCAACTCGTGGC 1025  
|||||

Db 200 CTCTACTGCAGGCCAAGATCGTGGGCTCTACGTCGCGCGCGCAACTCGCGGCG 259  
|||||  
QY 1026 CAGACCTACAGTCTGTGCTGTGTGTGTAGTGTGGCGAGGCC 1064  
|||||  
Db 260 CAGACGCTCAGTGGCGCTGAGCTGCACGTGAAGAGGCC 298  
|||||

RESULT 15  
AW170791  
LOCUS  
DEFINITION AW170791 407 bp mRNA linear EST 12-NOV-1999  
xj25b01.x1 NCI CGAP Ut2 Homo sapiens cDNA clone IMAGE:2658217 3',  
similar to TR:Q9ZA63 Q9ZA63 HYPOTHETICAL 30.2 KD PROTEIN. ;, mRNA  
sequence.

ACCESSION  
VERSION AW170791  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 407)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)

Possible reversed clone: similarity on wrong strand  
Seq primer: -40UP from gibco

High quality sequence stop: 236.

FEATURES  
Location/Qualifiers

1..407  
/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:2658217"

/tissue\_type="moderately-differentiated endometrial  
adenocarcinoma, 3 pooled tumors"

/lab\_host="DH10B"

/clone\_lib="NCI CGAP\_Ut2"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.85 kb. Life Technologies catalog #:  
11539-012"

ORIGIN

Query Match 10.5%; Score 112.8; DB 10; Length 407;  
Best Local Similarity 56.8%; Pred. No. 7.2e-06;  
Matches 227; Conservative 0; Mismatches 172; Indels 1; Gaps 1;

QY 638 GCTGGCGGCTGCAGCGCGCGCGCTAGTGTGTGAGCTCGGACCGCGGACCGCGC 697  
|||||

Db 5 GCTGGAGCGCGCGCGCTCCGACCGCGAGCTGCAGCGCGCGCGCTCCATCGCC 64  
|||||

QY 698 CAGCCGCGCGCGGACCTCCACGCGCGCGCTCTCTGGCGCACCTGCAGCGCGCGCGAGGC 757  
|||||

Db 65 GGGGTCCGGGAGGGGGCGCGCGCTCTCTACCGGGGCGCGCGCGCGCGGACC 124  
|||||

QY 758 TATGCGCGCGGAGGGGGCGCGCGCGCTCA-CCGCCAGACCGCGGACCGGACCTGCACCG 816  
|||||

Db 125 CCACAGGAGGCGCGCGCGCTGGTGTGTGAGCGCGCTCAAGTGCAGCGCTTAAGACCTCTGGG 184  
|||||

QY 817 TGACTGAAGCAAGCAGCGCGCGCTCAGCTGTACGTGCTACGCGCGGACCGCAAGCCGAG 876  
|||||

Db	185	TGAACGAGGGCAAGCACCCCAAGTTCCGCTGCTACGTGATGGCAAGCCCGAGCCGAGA	244
Qy	877	CGGTGTGGAAGAGACGCGCCAGCTGGTGACCGAGGGCCGCGCCACGTGGTGTACGAGG	936
Db	245	TCGAATGTCACGTGGAGGGCCCGCGCTGCTCCCGACCCCGCCGCTCATGTACCGCG	304
Qy	937	ACGCGCAGGAGAACTTTCGTGCTCAAGATCCTCTTCTGCAAGCAGTCGGACCGCGGCTCT	996
Db	305	ACCGCGACGCGGCTTCGTGCTCAAGGTGCTTTACTGCCAGGCCCCAGGATCGTGGGCTCT	364
Qy	997	ACACCTGCACGGGTCGAACTCGTGGGCCAGACCTACAG	1036
Db	365	ACGTTTGGCGCGCGCAACTCGCGGGCCAGACGCTCAG	404

Search completed: September 19, 2004, 22:07:21  
Job time : 3125.75 secs

OM nucleic - nucleic search, using sw model  
Run on: September 19, 2004, 14:16:08 ; Search time 3959.71 Seconds  
(without alignments)  
10628.572 Million cell updates/sec  
Title: US-10-077-130-4\_COPY\_23150\_24120  
Perfect score: 971  
Sequence: 1 cagcggcgccgctggccg.....ggcgcgtaaaaagtctaga 971  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.hcg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.ets.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.to.\*  
27: em.ats.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.vrt.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	946.4	97.5	4175	6	AX039409	AX039409 Sequence
2	926.8	95.4	5207	6	AR255867	AR255867 Sequence
3	926.8	95.4	5207	6	AX430856	AX430856 Sequence
4	926.8	95.4	5207	6	AX430858	AX430858 Sequence
5	911.4	93.9	3225	6	AX642969	AX642969 Sequence
6	911.4	93.9	5454	6	AX504255	AX504255 Sequence
7	906.4	93.3	4936	6	AX207401	AX207401 Sequence
8	905.4	93.2	5007	6	AX039410	AX039410 Sequence
9	905.4	93.2	7928	6	AX039412	AX039412 Sequence
10	884.4	91.1	4041	9	AB046859	AB046859 Homo sapi
11	631.8	65.1	1183	10	BC046431	BC046431 Mus muscu
12	262	27.0	62164	9	AL670729	AL670729 Human DNA
13	229.4	23.6	164766	2	AC026657	AC026657 Homo sapi
14	229.4	23.6	174612	2	AC023889	AC023889 Homo sapi
15	223.4	23.0	225	6	AX309418	AX309418 Sequence
16	207	21.3	174612	2	AC023889	AC023889 Homo sapi
17	206	21.2	7710	6	AX056397	AX056397 Sequence
18	204.4	21.1	7789	6	AX746187	AX746187 Sequence
19	198.8	20.5	4357	10	AK122488	AK122488 Mus muscu
20	189.4	19.5	10801	10	AF215896	AF215896 Mus muscu
21	179.2	18.5	6726	9	AB037718	AB037718 Homo sapi
22	176.6	18.2	260998	2	AC099089	AC099089 Rattus no
23	173.6	17.9	103129	10	AL662809	AL662809 Mouse DNA
24	173.6	17.9	258622	2	AC020877	AC020877 Mus muscu
25	126.8	13.1	182	6	AX913866	AX913866 Sequence
26	126.8	13.1	182	6	BD049399	BD049399 Sequence
27	113.8	11.7	1365	12	AX335738	AX335738 Synthetic
28	113.8	11.7	2055	9	AB022341	AB022341 Homo sapi
29	113.8	11.7	2079	6	BD127287	BD127287 Primer fo
30	113.8	11.7	2079	9	AK074759	AK074759 Homo sapi
31	113.8	11.7	2118	9	AK097643	AK097643 Homo sapi
32	113.8	11.7	2224	6	AX880020	AX880020 Sequence
33	113.8	11.7	2224	6	BD012208	BD012208 Novel gen
34	113.8	11.7	2224	6	BD158150	BD158150 Primer fo
35	113.8	11.7	2224	9	AK027590	AK027590 Homo sapi
36	112.2	11.6	2105	6	BD176607	BD176607 Method of
37	112.2	11.6	2105	9	AB007144	AB007144 Homo sapi
38	112.2	11.6	2132	6	AR076189	AR076189 Sequence
39	112.2	11.6	2132	6	AR124102	AR124102 Sequence
40	112.2	11.6	2132	6	E23384	E23384 DNA encodin
41	105.8	10.9	1719	8	AK109754	AK109754 Oryza sat
42	99.2	10.2	2731	8	AK110172	AK110172 Oryza sat
43	97	10.0	3870	5	CHESKELMUS	M81787 Gallus dome
44	96	9.9	1560	10	BC062076	BC062076 Rattus no
45	96	9.9	2903	4	RAEMLCKA	J05194 Rabbit myos

ALIGNMENTS

RESULT 1  
AX039409 AX039409 4175 bp DNA linear PAT 18-NOV-2000  
LOCUS Sequence 2 from Patent WO0063381.  
DEFINITION AX039409  
ACCESSION AX039409  
VERSION AX039409.1 GI:11229476  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Zeng W., Stanton L. and Kong H.  
TITLE Mammalian protein with putative function in signal transduction  
JOURNAL Patent: WO 0063381-A 2 26-Oct-2000;

SCIOS INC. (US)	
FEATURES	Location/Qualifiers
source	1. 4175
	/organism="Homo sapiens"
	/mol_type="unassigned DNA"
	/db_xref="taxon:9606"
CDS	<1..4056
	/note="unnamed protein product"
	/codon_start=1
	/protein_id="CAC16624.1"
	/db_xref="GI:11229477"
	/db_xref="REMBL: CAC16624"
	/translation="GVQFSQYGSPEFVSPHIIQONPVSEASDIWAGVIVSLTCS
	PFAGESRATLNLVLEGRVSWSPMAAHLDEDAKDFIKATLQAPQPSAAQCLSHP
	WFLKSPAEAEHFTIKQLKFLARWQSLKYSILVMRSIPELLRPDPSPSLG
	VARHLCDRTGSSSSSSONELAPFAKSLPSPVTHSPLLHPRGFPLPSALPDEE
	AEASERTEAPAPASPEGAPPAQOCVPHSVIRSLFYHAGSEPEHGALAPGER
	HPARRHLKGYIAGALPGLRPLMESHVLEBEAEEOATLLAKAPSPETALRLPA
	SGTHLAPGHSLSHSDPSRPSSEAGEAQLPSPASGAPIRDMGHQGSQKLPS
	TGHPHQAQPERPSPDFWGPAPFCHPKQGSAPQEGCSPPHVAAPCPGSPFPFGCK
	EAPLVPSFVLQGPAPAPAKASPLDLSKMGPDLSLPGRPKPGQCSFSGSASSS
	SOVSLRVSQVGTFRPGSLDARGMTQEAEDLSDSPTLQRPQEOATWKFSLQGRG
	GYSLPVGVTGFAFGAGAGMLGQGPWARIATAVQSBEEBEAEASQESRQBAR
	AEASERTEAPAPASPEGAPPAQOCVPHSVIRSLFYHAGSEPEHGALAPGER
	SEVPAPVNLSDLYDIKYLPEFEMI FRKVPKSAQPERPPSWAEELAEFPPTMPWG
	ELGHAGLITEESDDVALLAEAVGRKKNSSPSRSLFHFGRHLPLDEPAELGR
	ERVKASVEHSIRILKRGPEGLEKEGPPKPKGLASFLSLGKSWDRAPTLRELSDET
	VVLQGVTLACQVAAQAAQATWSKDGAPLESSESRVLSATLNFQLLTLVVAEDL
	GVYTCVSNALGVTITGVLKAEKPPSSPCDITGEVYADGVLLVWKPVSFVPTVI
	VOCLEGGSTWTLASDIFDCYLTSLSRGTYTFTACVSKAGMPSPPSPQVLLIG
	GPHSLAEESSQGSAPLPSTKTFATQTOIORGRFSVRQKWEKASGRALAAK11PY
	HPKDTAVLRVEYALKLRHPHLAQLHAIVLSPHLVLLLELCSPELLPLCLAESRY
	SESEVDKYLWQMSATQYLNQHLHLDLASENNIITEYNLLKVDLGNQASLSQKV
	LPSDFKDLTEMAPLELLEGGAGVFPDIDIAIGVTAFTILMSAEYFVPSSEGRDLQRL
	RKGLVRLSRCAVSGGAVAFRLSTLCAQWPRPCASSCLQCQPLWTEEGSPRPAPV
	TFPTARLVFVRNEKRRALLYKRNLAQVR"
ORIGIN	
Query Match	97.5%; Score 946.4; DB 6; Length 4175;
Best Local Similarity	99.9%; Pred. No. 2.1e-140;
Matches	947; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	1 CAGCGGGGGCGGTGGCGGCAAGATCATCCCTACCAACCCCAAGGACAGACAGT 60
Db	3228 CAGCGGGGGCGGTGGCGGCAAGATCATCCCTACCAACCCCAAGGACAGACAGT 3287
QY	61 GCTGGCGGAATACAGAGGCGCTCAGGGCGCTCGGCACCCGACCTGCGCCAGCTGCACGC 120
Db	3288 GCTGGCGGAATACAGAGGCGCTCAGGGCGCTCGGCACCCGACCTGCGCCAGCTGCACGC 3347
QY	121 AGCTTACCTCAGCCCCCGGACCTGGTCTCATCTTGGAGCTGTGCTCTGGGCGCGAGCT 180
Db	3348 AGCTTACCTCAGCCCCCGGACCTGGTCTCATCTTGGAGCTGTGCTCTGGGCGCGAGCT 3407
QY	181 GCTCCCTGCTGGCGGAGAGGGCGCTCTACTAGAAATCCGAGTGAAGGACTACCTGTG 240
Db	3408 GCTCCCTGCTGGCGGAGAGGGCGCTCTACTAGAAATCCGAGTGAAGGACTACCTGTG 3467
QY	241 CGAGATGTTGAGTGCACCCAGTACCTGACCAACAGCACATCTCTGACCTGCACCTGAG 300
Db	3468 CGAGATGTTGAGTGCACCCAGTACCTGACCAACAGCACATCTCTGACCTGCACCTGAG 3527
QY	301 GTCGAGAACATGATCATCCGAATAACCTCTCAAGTCTGAGCTGTGACCTGGGCAATGC 360
Db	3528 GTCGAGAACATGATCATCCGAATAACCTCTCAAGTCTGAGCTGTGACCTGGGCAATGC 3587
QY	361 ACAGAGCTCAGCAGGAGAGGTGCTGCTCTCAGACAGAGTTTCAAGGACTACCTAGAGAC 420
Db	3588 ACAGAGCTCAGCAGGAGAGAGGTGCTGCTCTCAGACAGAGTTTCAAGGACTACCTAGAGAC 3647
QY	421 CATGGCTCCAGAGCTCTGGAGGCGCCAGGGGCGTGTTCACAGACAGACATCTGGGCCAT 480
Db	3648 CATGGCTCCAGAGCTCTGGAGGCGCCAGGGGCGTGTTCACAGACAGACATCTGGGCCAT 3707

QY	481	CGGTGTGACAGCCTTTCATCATGCTGAGCGCCGAGTAGTACCCGGTGTAGCAGGAGGTTGCACG	540
Db	3708	CGGTGTGACAGCCTTTCATCATGCTGAGCGCCGAGTAGTACCCGGTGTAGCAGGAGGTTGCACG	3767
QY	541	CGACTTGCAGAGAGGACTGCGCAAGGGGCTGGTCCGGCTGAGCCGCTGCTAGCCGGGGCT	600
Db	3768	CGACTTGCAGAGAGGACTGCGCAAGGGGCTGGTCCGGCTGAGCCGCTGCTAGCCGGGGCT	3827
QY	601	GTCGGGGGGCGCGTGGCCTTCTGTGCGACAGCATCTGTGGGCCAGCCCTTGGGGCGCGCC	660
Db	3828	GTCGGGGGGCGCGTGGCCTTCTGTGCGACAGCATCTGTGTGCGCCAGCCCTTGGGGCGCGCC	3887
QY	661	CTGCGCGCTCCAGGCTGCCTGTGACGTGCCCGTGGCTTAAACAGAGGAGGGCCCGGCTGTTCGGG	720
Db	3888	CTGCGCGCTCCAGGCTGCCTGTGACGTGCCCGTGGCTTAAACAGAGGAGGGCCCGGCTGTTCGGG	3947
QY	721	GCCCGCGCCCGTGCACCTTCCCTACCGCGCGGCGTGCSCGTCTTCGTGCGCAATATGCGAGNA	780
Db	3948	GCCCGCGCCCGTGCACCTTCCCTACCGCGCGGCGTGCSCGTCTTCGTGCGCAATATGCGAGNA	4007
QY	781	GAGACGCGCGCTGCTGTACAAAGAGGCACAACCTGGCCCGCAGGTGCGTGAAGGGTTCGCCCCG	840
Db	4008	GAGACGCGCGCTGCTGTACAAAGAGGCACAACCTGGCCCGCAGGTGCGTGAAGGGTTCGCCCCG	4067
QY	841	GCCACACCTTGGTCTCCCGCTGGGGTGCCTGTGCAGACGCGCCCAATAAAAAAGCCCGACG	900
Db	4068	GCCACACCTTGGTCTCCCGCTGGGGTGCCTGTGCAGACGCGCCCAATAAAAAAGCCCGACG	4127
QY	901	CGGCGCAGAAAAAIAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 948	
Db	4128	CGGCGCAGAAAAAIAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4175	
RESULT 2			
LOCUS	AR255867	5207 bp	DNA linear PAT 20-DEC-2002
DEFINITION	Sequence 1 from patent US 6482624.		
ACCESSION	AR255867		
VERSION	AR255867.1	GI:27305065	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 5207)		
AUTHORS	Wei, M.-H., Ketchum, K.A., Di Francesco, V. and Beasley, E.M.		
TITLE	Isolated human kinase proteins, nucleic acid molecules encoding human kinase proteins, and uses thereof		
JOURNAL	Patent: US 6482624-A 1 19-NOV-2002;		
FEATURES	Location/Qualifiers		
source	1. 5207		
	/organism="unknown"		
	/mol_type="genomic DNA"		
ORIGIN			
Query Match	95.4%;	Score 926.8; DB 6;	Length 5207;
Best Local Similarity	99.8%;	Pred. No. 2.6e-137;	
Matches	928; Conservative 0;	Mismatches 2;	Indels 0; Gaps 0;
QY	1	CAGCGGGCGGCGCTGGCCCGCAAGATCATCCCTACCAACCCCAAGGACAGACAGT	600
Db	4278	CAGCGGGCGGCGCTGGCCCGCAAGATCATCCCTACCAACCCCAAGGACAGACAGT	4337
QY	61	GCTCGCGGAATACAGAGGCGCTCAAGGGCGCTGCGCCACCCGACCTGCGCCAGCTGCACGC	120
Db	4338	GCTCGCGGAATACAGAGGCGCTCAAGGGCGCTGCGCCACCCGACCTGCGCCAGCTGCACGC	4397
QY	121	AGCTTACCTCAGCCCCCGGACCTGGTGTGCTCATPTTTGGAGCTGTGCTTGGGCGCGAGCT	180
Db	4398	AGCTTACCTCAGCCCCCGGACCTGGTGTGCTCATPTTTGGAGCTGTGCTTGGGCGCGAGCT	4457
QY	181	GCTCCCTGCTGCGCGAGAGGGCGCTCTACTCAGAAATCCGAGGTGAAGGACTACTCTGTG	240



ORIGIN	Query Match	Best Local Similarity	Score	926.8;	DB 6;	Length	5207;
	Matches	Conservative	0;	Mismatches	2;	Indels	0;
	0;	Gaps	0;				
QY	1	CAAGCGGCGGCGCTGGCCGCAAGATCATCCCTACCA	95.4%;	99.8%;	Pred. No. 2.6e-137;		
Db	4278	CAAGCGGCGGCGCTGGCCGCAAGATCATCCCTACCA					
QY	61	GTTGGCGCAATACGAGGCGCTCAAGGCGCTGGCC					
Db	4338	GTTGGCGCAATACGAGGCGCTCAAGGCGCTGGCC					
QY	121	AGCTACCTCAGCCCGCGCAGCTGGTGCTCATCTT					
Db	4398	AGCTACCTCAGCCCGCGCAGCTGGTGCTCATCTT					
QY	181	GTCCTCCCTGCCCTGGCGAGGCGCTCTACTTCAG					
Db	4458	GTCCTCCCTGCCCTGGCGAGGCGCTCTACTTCAG					
QY	241	CGAGATGTTAGTGGCCAGCCAGTACCTGCA					
Db	4518	CGAGATGTTAGTGGCCAGCCAGTACCTGCA					
QY	301	GTCGAGAACATGATCATCACCGAATAACA					
Db	4578	GTCGAGAACATGATCATCACCGAATAACA					
QY	361	ACAGAGCCTCAGCAGGAGAGAGTGTCTCCCT					
Db	4638	ACAGAGCCTCAGCAGGAGAGAGTGTCTCCCT					
QY	421	CATGCTCCAGAGCTCCTGGAGGGCCAGGGGGCT					
Db	4698	CATGCTCCAGAGCTCCTGGAGGGCCAGGGGGCT					
QY	481	CGGTGTGACAGCTTCATCATGCTGAGCGCGAG					
Db	4758	CGGTGTGACAGCTTCATCATGCTGAGCGCGAG					
QY	541	CGACTTCGAGAGAGACTCGCAAGGGGCTGGT					
Db	4818	CGACTTCGAGAGAGACTCGCAAGGGGCTGGT					
QY	601	GTCCGGGGGGCGGTGGGCTTCCTCGCAGACACT					
Db	4878	GTCCGGGGGGCGGTGGGCTTCCTCGCAGACACT					
QY	661	CTGCGCTCCAGCTGCTCGAGTGCCTGGCTAAC					
Db	4938	CTGCGCTCCAGCTGCTCGAGTGCCTGGCTAAC					
QY	721	GCCTCGCGCCGTGACCTTCCTACCGCGCGCT					
Db	4998	GCCTCGCGCCGTGACCTTCCTACCGCGCGCT					
QY	781	GAGACGCGCTGTGTGTACAGAGGCAACCTGG					
Db	5058	GAGACGCGCTGTGTGTACAGAGGCAACCTGG					
QY	841	GCACACCTTTGGTCTCCCGCTGGGGTGCCTGC					
Db	5118	GCACACCTTTGGTCTCCCGCTGGGGTGCCTGC					
QY	901	CGGGCGAGAAAAA					
Db	5178	CGGGCGAGAAAAA					

RESULT 4  
AX430858





```

VERSION      AX207401.1  GI:15395213
KEYWORDS
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    1
AUTHORS      Ploeman, G., Whyte, D., Manning, G., Sudarsanam, S. and Martinez, R.
TITLE        Human protein kinases and protein kinase-like enzymes
JOURNAL      Patent: WO 0155356-A 14 02-AUG-2001;
              Sugan, Inc. (US)
FEATURES     Location/Qualifiers
              1..4936
              /organism="Homo sapiens"
              /mol_type="unassigned DNA"
              /db_xref="taxon:9606"
ORIGIN
Query Match      93.3%; Score 906.4; DB 6; Length 4936;
Best Local Similarity 99.9%; Pred. No. 4.4e-134;
Matches 907; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGGGGGCGGTGGCGGCAAGATCATCCCTACGACCCCAAGGACAGCAGT 60
DB 4029 CAGCGGGGGCGGTGGCGGCAAGATCATCCCTACGACCCCAAGGACAGCAGT 4088

QY 61 GCTGCGGAATACAGAGCCCTCAAGGGCTGCGCCACCCGACCTGCGCCAGTGCACGC 120
DB 4089 GCTGCGGAATACAGAGCCCTCAAGGGCTGCGCCACCCGACCTGCGCCAGTGCACGC 4148

QY 121 AGCTACTCAGCCCCCGGACCTGCTCATCTTGAGCTGTGCTCTGGCCCGGAGCT 180
DB 4149 AGCTACTCAGCCCCCGGACCTGCTCATCTTGAGCTGTGCTCTGGCCCGGAGCT 4208

QY 181 GCTCCCTGCTGCGGAGAGGCTCTCTACTAGATCCGAGTGAAGACTACCTGTG 240
DB 4209 GCTCCCTGCTGCGGAGAGGCTCTCTACTAGATCCGAGTGAAGACTACCTGTG 4268

QY 241 CGAGATGTGAGTGCCACCCAGTAGTACCTGCACCAACGACACATCCTGCACCTGGACCTGAG 300
DB 4269 CGAGATGTGAGTGCCACCCAGTAGTACCTGCACCAACGACACATCCTGCACCTGGACCTGAG 4328

QY 301 GTCCGAGAACATGATCATACCGAATACACCTGCTCAAGGTGTGACCTGGGCAATGC 360
DB 4329 GTCCGAGAACATGATCATACCGAATACACCTGCTCAAGGTGTGACCTGGGCAATGC 4388

QY 361 ACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTCAAGGACTACCTAGAGAC 420
DB 4389 ACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTCAAGGACTACCTAGAGAC 4448

QY 421 CATGGCTCCAGAGCTCTCGAGAGGCCAAGGGGCTGTTCACAGACAGACATCTGGGCGCAT 480
DB 4449 CATGGCTCCAGAGCTCTCGAGAGGCCAAGGGGCTGTTCACAGACAGACATCTGGGCGCAT 4508

QY 481 CGGTGTGACAGCTTTCATCTGCTGAGCGCGAGTACCCGCTGAGCAGCAGGAGGTGCACG 540
DB 4509 CGGTGTGACAGCTTTCATCTGCTGAGCGCGAGTACCCGCTGAGCAGCAGGAGGTGCACG 4568

QY 541 CGACTCGAGAGAGGACTCGCAAGGGGCTGGTCCGCTGAGCGGCTGTGACCGGGGCT 600
DB 4569 CGACTCGAGAGAGGACTCGCAAGGGGCTGGTCCGCTGAGCGGCTGTGACCGGGGCT 4628

QY 601 GTCCGGGGGCGCTGCTTCCCTGCGACAGCTCTGTGCGCCGACCGCTGGGCGCGGC 660
DB 4629 GTCCGGGGGCGCTGCTTCCCTGCGACAGCTCTGTGCGCCGACCGCTGGGCGCGGC 4688

QY 661 CTGCGGCTTCAGAGCTGCTGAGTCCCGCTGGCTAACAGAGAGGGCGCCGCTGTTCGCG 720
DB 4689 CTGCGGCTTCAGAGCTGCTGAGTCCCGCTGGCTAACAGAGAGGGCGCCGCTGTTCGCG 4748

QY 721 GCCCGCGCGCTGACCTTCCCTAGCGCGGCTGGGCTCTTCTGCGCAATCGCGAGAA 780
DB 4749 GCCCGCGCGCTGACCTTCCCTAGCGCGGCTGGGCTCTTCTGCGCAATCGCGAGAA 4808
```

```

QY 781 GAGACGCGCGCTGCTGTACAAGAGGACCAACCTGCGCCAGGTGCGCTGAGGGTGCCTCCCG 840
DB 4809 GAGACGCGCGCTGCTGTACAAGAGGACCAACCTGCGCCAGGTGCGCTGAGGGTGCCTCCCG 4868

QY 841 GCACACACCTTGGTCTCCCGCTGGGGTCTGCTGAGACGCGCCCAATAAAGACGCCAGC 900
DB 4869 GCACACACCTTGGTCTCCCGCTGGGGTCTGCTGAGACGCGCCCAATAAAGACGCCAGC 4928

QY 901 CGGCGGAG 908
DB 4929 CGGCGGAG 4936

RESULT 8
AX039410
LOCUS      AX039410
DEFINITION Sequence 3 from Patent WO0063381.
ACCESSION AX039410
VERSION    AX039410.1  GI:11229478
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1
            Zeng, W., Stanton, L. and Kong, H.
            Mammalian protein with putative function in signal transduction
            Patent: WO 0063381-A 3 26-OCT-2000;
            SCIOS INC. (US)
FEATURES   Location/Qualifiers
            1..5007
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            97..4929
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CAC16625.1"
            /db_xref="GI:11229479"
            /translation="MQVTIEDVQAQTGTAQFEALIEGDPQPSVTWVKDSVQLVDSTRE
            LSQEQGTTYSLVLRHVASKDAGVYTCIAQTGQVLCXAEALLVLGDNBDESKQSH
            RKUHSFYVEKEEIGRGVFGKRVKQKNGKILCAKPIRSTRAQATREDDILAA
            LSHPLVTGLDQFETRTLLILSEELDLRYKVVTEAEVKYIQQLVEGHU
            YLHSHVLDLIDIKFENILMVHAREDIKIDFGAQTITPAELQFQYQSGPVSPI
            IQQNPVSEASDIWAGVISYLSLTCSSPFAGSDRATLNLVLEGRVSWSPMAHLE
            DAKDPIKATLQAPQAPSAACLSHFVFLKSMFAEAFINTKQLKFLARSQWS
            LMSYKSLIVMSIPELLRPPDSPSLGVARHLCRDTGSGSSSSSSSNELAPPARAKS
            LPSPVTHPLHPRGFLRPSASLPEAEASERSTEAPAPASPEGAGPAAQCVPR
            HSVIRSFYHQAGSPERHAPGRHRRHRRHLLKGYTAGALPGLRPLMEHVR
            EEEAREQATLLAKAPSFETALRPSAGTHLAPGSHSHUEHDSPTTPRSEACGBA
            QRLPSAGGAPIRDMGHPQSKQLPSTGHGPTAQPERSPDPSPWQCPAPFCHPKQM
            SAQEGCSPPHAPVPCPPGPPGCKEAPLVPSPFLGQAPAPAKASPLDQSM
            GPDISLFGPKPGPCSPGSAQSSQVSLRVGSQVTEGTEPSPDIAEGMTQEAR
            DLSSTPLQRPQCATMRKESLAGRGYAGVAGYGFAGGAGGVLGGGPMWART
            WAWQSEEEEEEARASQSEEQEARESPLPQVSARVPYENGRATRSAPETPE
            DIGQVSLVIRDLSDAEADTISDISEVDFAVLNLSLDYIKIYFFEFMFKRKP
            SAQPEPSPMAEEELAFPEFTWFWPGLGHAGLEITSEEDVALLAAAVGRKR
            WSSPSRSLFHPFGRHLPLDEPAGLRRKVASVEHISRLIKRPEGLEKEGPPRKP
            GLASPRSLGKSWDRAPTFRLSDTETVLQSVTLACQSAQAPAAQATSKDGAPLE
            SSSVLISATLNFQLLTILVVAEDILVYCSVSNALGTVTTTGLRKARKEPSSPC
            PDIGVADGVLVWKVPSXGPTVYVQCSLEGSSWTTLASDIFDCYLTSKUSRG
            TYTRTACVSKAGMPYSSPSEOVLLGAPSHLASEEESQGRSAQPLSTTFAQTOI
            QGRFSPVQRCWAKSARNAKLIIPHPKDTAVLREYALKGLRHLHAQLHAAVL
            SPRHLVLELSCGPELLPCLAAERASYSESEVLDLWQMLSATQYLHNQHLHLDLRS
            ENMIIETYNLKVVDLGNLSQEKVLPDQKFDYLETWAPLEEGCGAVPQTDIWA
            IGVTAIFMLSAEYVPSSEGARDLGRKGLVRSRCYAGLSGGAFLRSTLCAQPW
            GRPCASSCLQCPWLTTEGPACSRPAPVFTFARLVRVFRNRKRALLYKRNLAQVR
            "
ORIGIN
Query Match      93.2%; Score 905.4; DB 6; Length 5007;
```

Best Local Similarity 99.9%; Pred. No. 6.3e-134; Matches 906; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
QY	1	CAGCGGGCGGCGCTGGCGGCAAGATCATCCCTACACCCCAAGGACAGACAGCAGT	60						
DB	4101	CAGCGGGCGGCGCTGGCGGCAAGATCATCCCTACACCCCAAGGACAGACAGCAGT	4160						
QY	61	GCTGCGCAATACGAGGCGCTCAAGGGCTCGGCAACCGGACCTGCGCCAGTGCACG	120						
DB	4161	GCTGCGCAATACGAGGCGCTCAAGGGCTCGGCAACCGGACCTGCGCCAGTGCACG	4220						
QY	121	AGCCTACCTACGCGCCGCGACCTGCTCATCTTGAGCTGCTCTGGGCGCGAGCT	180						
DB	4221	AGCCTACCTACGCGCCGCGACCTGCTCATCTTGAGCTGCTCTGGGCGCGAGCT	4280						
QY	181	GCTCCCTGCTGCGCGAGAGGGCTCTCTACTCAGATCCGAGGTGAAGGACTACCTGTG	240						
DB	4281	GCTCCCTGCTGCGCGAGAGGGCTCTCTACTCAGATCCGAGGTGAAGGACTACCTGTG	4340						
QY	241	GCAGATGTTGAGTGCACCGCAGTACTGTCACACCGACATCTGTCACCTGGACCTGAG	300						
DB	4341	GCAGATGTTGAGTGCACCGCAGTACTGTCACACCGACATCTGTCACCTGGACCTGAG	4400						
QY	301	GTCGAGAACATGATCATCACCAGATACAACTGCTCAAGGTCTGACCTGGCAATGC	360						
DB	4401	GTCGAGAACATGATCATCACCAGATACAACTGCTCAAGGTCTGACCTGGCAATGC	4460						
QY	361	ACAGAGCCTCAGCAGGAGAGGTGCTGCTCCCTCAGACAGTTCAGGACTACCTAGAGAC	420						
DB	4461	ACAGAGCCTCAGCAGGAGAGGTGCTGCTCCCTCAGACAGTTCAGGACTACCTAGAGAC	4520						
QY	421	CATGCTCCAGAGTCTCTGAGGGGCGAGGGGCTGTTCCACAGACAGACATCTGGGCGAT	480						
DB	4521	CATGCTCCAGAGTCTCTGAGGGGCGAGGGGCTGTTCCACAGACAGACATCTGGGCGAT	4580						
QY	481	CGGTGTGACAGCCTTCATCATCTGAGCGCGAGTACCCGTGAGCAGCGAGGTGACG	540						
DB	4581	CGGTGTGACAGCCTTCATCATCTGAGCGCGAGTACCCGTGAGCAGCGAGGTGACG	4640						
QY	541	CGACTGACAGAGAGTCTGCGAAGGGCTGCTCGGCTGAGCCGCTGCTACCGGGGCT	600						
DB	4641	CGACTGACAGAGAGTCTGCGAAGGGCTGCTCGGCTGAGCCGCTGCTACCGGGGCT	4700						
QY	601	GTCGCGGGCGCGCTGCGCTTCTGCGGAGCAGTCTGTCGCCCGCAGCCCTGGGGCGCGGC	660						
DB	4701	GTCGCGGGCGCGCTGCGCTTCTGCGGAGCAGTCTGTCGCCCGCAGCCCTGGGGCGCGGC	4760						
QY	661	CTGCGCTGCGAGTCTGCGAGTGGCGGCTGCTGAGAGAGGGCGCGGCTGCTGCGG	720						
DB	4761	CTGCGCTGCGAGTCTGCGAGTGGCGGCTGCTGAGAGAGGGCGCGGCTGCTGCGG	4820						
QY	721	GCCGCGCGCGTGACCTTCCCTACCGCGCGCTGCGGCTGCTGCGGCAATCGCGAGAA	780						
DB	4821	GCCGCGCGCGTGACCTTCCCTACCGCGCGCTGCGGCTGCTGCGGCAATCGCGAGAA	4880						
QY	781	GAGACGCGCGTGTGTACAAGAGCACAACCTGCGCCAGTGCCTGAGGCTGCGCCCG	840						
DB	4881	GAGACGCGCGTGTGTACAAGAGGACAACCTGCGCCAGTGCCTGAGGCTGCGCCCG	4940						
QY	841	GCCACACCTGCTGCTCCCGCTGGGGTCTGCTGAGAGCGCGCAATAAAGACCGCCAGC	900						
DB	4941	GCCACACCTGCTGCTCCCGCTGGGGTCTGCTGAGAGCGCGCAATAAAGACCGACAGC	5000						
QY	901	CGGGCGA 907							
DB	5001	CGGGCGA 5007							
RESULT 9									
AX039412									
LOCUS									
DEFINITION									
SEQUENCE 5 from Patent WO0063381.									
ACCESSION									
AX039412									
linear									
DNA									
7928 bp									
PAT 18-NOV-2000									

AX039412.1	GI:11229480								
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
JOURNAL									
FEATURES									
source									
	1..7928								
	/organism="Homo sapiens"								
	/mol_type="unassigned DNA"								
	/db_xref="taxon:9606"								
	60..7850								
CDS									
	/codon_start=1								
	/protein_id="CAC16626.1"								
	/db_xref="GI:11229481"								
	/db_xref="RENTREMBL:CAC16626"								
	/translation="MLRFTPKVKKSSITFSVKVEGRVPTVHMLEEAEGRGLVMI								
	GPDTGYTVASSAQHSLVLDVGRHQCTYTTCIASNAAGQALCSASLHVSGLPVKEE								
	QEKVREALSTFLOQTTOAISAGLETAFADLGGORKEEPLAAKEALGHSLAEVGT								
	EEFLQKLSQITEMUSAKITQAKLOVPGSDSDSKTPSASPRHSRSPSSSIQSSSS								
	ESEDGARGEIDFIYVVTADYLPQAEQDAILREGQVVEVLDAAHPLRWLVRTPTK								
	SPSQGWSPAYLDRLRLKSPWGAARAEPPFGAEVSEDEYKALSSSVIQELLSSEQ								
	AFVEELQFQSHLQHLERCPHVIAVAGQKAVIPRNVRDIGHFSSFIQELQCCDQ								
	DDVACFIKNQAAFEQYLEFLVRQVAAVVSTAIQEFYKKAEEALAGAEFSQPP								
	PLQHYLEQVQVQVQYQALLKELIRNARNRNCALLQAVYVSAALPQRAENKLVH								
	SIMENVPGTLEALGPRIQGHFIWEGAPGARMKPKHNRHVELFNHVLICPKRPS								
	RMENVYVFRNMKLSSIDLNDVGGDRAPEVQWEDSVKYLLOARTAIISKSV								
	KEICGQRLALPVWRPDPFEELADCTAELGETVKLACRVGTGTPKPVISWYDKGAV								
	QVDPHILIEDPDGCALDILSLTGQWQCMFAASAGNCSITGKILGVPPVPRV								
	KYRASPFEVEDAQFTIEGAPYQIRWYKDGALLTTGNKFTLSEPRSGLLVIR								
	AASKEDLGLYECELVNRLGSAASAEALIQSPMLQAEQCHREQLVAAVEDTLERAD								
	QSVTLVKRLGPKAPSTGDLTGPCPCPGAPALQETGQSPVPTGTEAPVPRV								
	POPLAHEGPEQPEALARAQWTVPIWEGAAPGAGTGLLDVHSHVVRTEORTY								
	TYQADITHTRAPSPMQVIEDVQAGTGAPEALIEGDPQSPVTKYKSVQLQSTR								
	LSQQEGTYSILVRHVASKADAGVITCLAQNTGGQVLCCKLELLVGGNEDPSEKQSH								
	RKLHSFVEKEEIRGVGFVKYVQHKNKILCAAKFPLNRSRTRAAQYRERDILAA								
	LSHPLVTGLDQFETRKTLILILECSBELLDRLYRKGVTEAVKVIQQLVEGLH								
	YHSHGVLHLDIKPNILMVHPAREDICDFGQANTIPAELOFSQSGPFFVSPPEI								
	IQONPVSEASDIWANGVLSLSLTCSPPAGESDRATLLNVLEGRVSSVMAAHLSE								
	DAKDFIKATLQAPQAPSAACLSHPWFLKSMPEAEAHFINTKLTLLARSWORS								
	LMYSKILLVRSIPELLRGPPSPSLGVARHLCRDTGSSSSSSSDHLELAPARAKS								
	LPSPVTHPLUHPGFLRPSASLPEEAEASERSTEAPAPSPASGAPFAAQGVPR								
	HVTRLSFYHQGESPEHGALAPGSRHRPARRHLLKGGYIAGLPGLURBPLMEHRVL								
	EEAAEEQATLLAKAPSETALRLPASGTHLAPGKSHLSHDSPTSPSSEACGEA								
	QRLPAPSGGAPIRDMHPQSKQLPSTGHPGTAQPRPSPDPSPWGPAPFHPKQK								
	SAPQEGCSHPAVAPCPGSPFPKCAPLVPSPPFLGQPOAPAPAPAPAPPLDSKM								
	GGDLSLPGRPKPCSPGSAOSSQVSSLRVGSQVTEGPGSLDAEGWTQEAR								
	LDSDSTPLQRPQEQATNRKFSLGGRGYAGVAGYGTAFGGDAGMLGQGPWMAIA								
	WAVSQSEEEAEARAESQSEEQEARAESPLQVSARPVFGVAGKAPINSPEPTWE								
	DIGQVSLVQIRLSDGAEAAATISLIDISEVDPAYNLSLDYIKYLPFFMIFRKYPK								
	SAQPPSPMAEEELAEFPETWMPGELGPHAGLEITESEEDVALLAAEAVGRRK								
	WSPSRSLFHPFGRHLPLDEPAELGRVRKASVEHISRIKLRGPEGLEKEGPPKPK								
	GLASFLSLGKSWDRAPFLRELSDETIVLGQSVTLACQVSAQAAQATWSKDGAPLE								
	SSSRVLISATLKNFOLLIVVVAEDLVGTVCSVNALGTVTTTGLVCKABRSPSSPC								
	PIGVIYADGVLLVWKPVESQPVYIYQCSLEGGSWTTLASDIEDCCCTSLKLRGG								
	TYTFRACVQKAGMGPYSPSEVLLGAPSHLASEESQGRSAQDPSTKTTFATQOI								
	QGRFSVVRQCWEKAGSALAAKIIPYHPKDTAVLRYEALKLGRHLHQAQHLHAYL								
	SPRHLVILLELCSGPELLCLAEARASYSEVKYLLWQLSATQYLHQLHLHLDRS								
	ENNIITEYNLLKVDLGNALQSQKVLPSDKFDYLETMAPLELLEGGAGVPTQDIWA								
	IGVTAFIMLSAEYPVSVSEGDQLQRLKGLVRLSRVAGLSGGAFAVRLTLCQWP								
	GRPCASSCLQCPWLTEEGPACSRPAPVTFPTARLVVRNREKRALLYKRNHLAQVR								
	"								

ORIGIN

Query Match 93.2%; Score 905.4; DB 6; Length 7928;  
Best Local Similarity 99.9%; Pred. No. 6e-134;  
Matches 906; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGGGCGGCTGGCGCCAGATCATCCCTACCAACCCCAAGGACAGACAGT 60  
 Db 7022 CAGCGGGCGGCGTGGCGCCAGATCATCCCTACCAACCCCAAGGACAGACAGT 7081  
 QY 61 GCTCGCGGAATACAGAGGCGCTCAAGGGCTCGCGCAACCCGACCTGGCCAGCTGCACGC 120  
 Db 7082 GCTCGCGGAATACAGAGGCGCTCAAGGGCTCGCGCAACCCGACCTGGCCAGCTGCACGC 7141  
 QY 121 AGCTTACCTCAGCCCCCGGACCTGGTCTCATCTTGGAGCTGTCTCTGGGCCCGAGCT 180  
 Db 7142 AGCTTACCTCAGCCCCCGGACCTGGTCTCATCTTGGAGCTGTCTCTGGGCCCGAGCT 7201  
 QY 181 GTCCTCCCTGCTGGCGGAGAGGCTCTCTACTCAGAAATCCGAGCTGAAGGACTACCTGTG 240  
 Db 7202 GTCCTCCCTGCTGGCGGAGAGGCTCTCTACTCAGAAATCCGAGCTGAAGGACTACCTGTG 7261  
 QY 241 GCAGATGTTAGTGCACCCAGTACCTGCACAACACGACACATCTCTGACCTGCACCTGAG 300  
 Db 7262 GCAGATGTTAGTGCACCCAGTACCTGCACAACACGACACATCTCTGACCTGCACCTGAG 7321  
 QY 301 GTCGAGAAATGATCATCAGCAATCAACCTCTCAAGTCTGAGCTGGGCAATGC 360  
 Db 7322 GTCGAGAAATGATCATCAGCAATCAACCTCTCAAGTCTGAGCTGGGCAATGC 7381  
 QY 361 ACAGAGCTCAGCGAGGAGAGTGTCTCTCAGCAAGTTCAGGACTACCTAGAGAC 420  
 Db 7382 ACAGAGCTCAGCGAGGAGAGTGTCTCTCAGCAAGTTCAGGACTACCTAGAGAC 7441  
 QY 421 CATGCTCCAGAGCTCTGGAGGCGCGAGGCGCTTCCACAGACACATCTGGGCCAT 480  
 Db 7442 CATGCTCCAGAGCTCTGGAGGCGCGAGGCGCTTCCACAGACACATCTGGGCCAT 7501  
 QY 481 CGGTGTGACGCTTCATCATGCTGAGCGCGAGTACCCGCTGAGCAGGAGGTGCACG 540  
 Db 7502 CGGTGTGACGCTTCATCATGCTGAGCGCGAGTACCCGCTGAGCAGGAGGTGCACG 7561  
 QY 541 GCACCTCAGAGAGAGCTGCGCAAGGGCTGGTCCGGCTGAGCGCTGCTACGGGGCT 600  
 Db 7562 GCACCTCAGAGAGAGCTGCGCAAGGGCTGGTCCGGCTGAGCGCTGCTACGGGGCT 7621  
 QY 601 GTCGGGGCGCGCTGGGCTTCTCTGCGCAGACCTCTGTGGCCAGCGCTGGGGCCCGGCC 660  
 Db 7622 GTCGGGGCGCGCTGGGCTTCTCTGCGCAGACCTCTGTGGCCAGCGCTGGGGCCCGGCC 7681  
 QY 661 GTGGCGCTCAGCTGCTGAGTGCCTGCTTAAAGAGAGGAGGCGCGGCTGTTCGCG 720  
 Db 7682 GTGGCGCTCAGCTGCTGAGTGCCTGCTTAAAGAGAGGAGGCGCGGCTGTTCGCG 7741  
 QY 721 GCCCGCGCTGACCTTCCCTACCGCGCGCTGCGCTCTTCTGTGGCAATCGCAGAA 780  
 Db 7742 GCCCGCGCTGACCTTCCCTACCGCGCGCTGCGCTCTTCTGTGGCAATCGCAGAA 7801  
 QY 781 GAGACGCGCTGCTGTAAAGAGGCAACCTGGCGCAGGTGCTGAGGGTGCAGG 840  
 Db 7802 GAGACGCGCTGCTGTAAAGAGGCAACCTGGCGCAGGTGCTGAGGGTGCAGG 7861  
 QY 841 GCCACACCTTGTCTCCCGCTGGGGTGTCTGCAGACGCGCAATAAAGAGCGCCAGC 900  
 Db 7862 GCCACACCTTGTCTCCCGCTGGGGTGTCTGCAGACGCGCAATAAAGAGCGCAGC 7921  
 QY 901 CGGGCGA 907  
 Db 7922 CGGGCGA 7928

## RESULT 10

LOCUS AB046859 4041 bp mRNA linear PRI 22-FEB-2001  
 DEFINITION Homo sapiens mRNA for KIAA1639 protein, partial cds.  
 ACCESSION AB046859  
 VERSION AB046859.1 GI:10047354  
 KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## source

## Location/Qualifiers

## 1. 4041

## /organism="Homo sapiens"

## /mol\_type="mRNA"

## /db\_xref="taxon:9606"

## /clone="fj06072"

## /note="vector:pBluescriptII SK plus"

## 1. 4041

## /gene="KIAA1639"

## /gene="KIAA1639"

## /note="start codon is not identified. fj06072 cDNA clone

## 2845 and 2901, and a 1-bp insertion after the position

## 1640 of the sequence of KIAA1639, respectively."

## /product="KIAA1639 protein"

## /protein\_id="BAB13465.1"

## /db\_xref="GI:10047355"

## /translation="AMGVISYLSLTCSPPFAGESDRATLNLVLEGRVSWSPMAHLS

## EDAKFIATLQRAFPQAPSAQCLSHPWFLKSNPAAEAFINTKQLKFLARSRWR

## SIMSYKSLVNRSPDLLRGPDPSPVARHLCDTGGSSSSSSSDNELAPFARAK

## SLPPSPVTHSLHPRGLRPSASLPBAAEASERSTAPAPAPAGAPAAQCVV

## RHSVRSFYHQAGSPHGLALPGRHPRARRHLKGGVYAGALPGLRPLMEHRV

## LEEARAREOATLLAKSPFETALRPLASGTHLAPGHSHSLHSDSPSTPRSSACCE

## AQRPLASGAPARDMGHPQSKQLRSTGGHPTAOPERSPDSPWQAPAPCPHQ

## GSAQPGCSHPFAPVAPCPFGSPGCKEAPLVFSPFLGQFPAPAPAKAPPLDK

## MGPGDISLPGRKPGPCSSPGSSAQSSQSSVLSRVSSQVGTFFGSLDAEGWTQA

## EDLSSTLPQRPOEQATMRKXSLGGRGVAGVAGYCTFAFGDAGMLGGGPMWAI

## AWAYSQSEEEQBARAESQEQEQAESPLQVGRAPVPEVGRAPTRSPPTPW

## EDIGVSLVQIRDLSDGAADTISLDSIVDPAYLNLSDLYDKYLPFFPMIERYVP

## KSAQPEPSPVAERELAFPRPTWPGELGPHAGLETTESEVDALLAEVGRKK

## KWSFSPSLFHPGRLPLDEPAGLRLERKVASVEHISRLKGRPGLEKEGPRKK

## PCLASFRUSGLKSWDRAPTFRELSDTVLQSVTLACQVSAQAPAAQATWSKQAPL

## ESSSRVLISATLKNFQTLILVVAEDLVGVTSVSNALGVTTTGVLRKAERPSPP

## CPDTEGVADVLLVWKFVSYGVTVIVQSLGGSWTILASIDPCCVLTSLKSG

## GTYTFRTACVSKAGMPYSPSEVLLGGPSHLASEBESQGRSAQPLSTFTAFQTC

## IQRGFSVVRQCEKASGRALAAIIPVHKDKTAVLRLEYEALKGLRPHLAQHLAA

## LSPRLVILLELCGPELLPCLAAASVSESEVDLWMLSATQYLHNOHLLHLDR

## SENMLITYNLLKVDLGNASLSQEKVLPSPDKDYLETMELLEGGOGAVPTDIMA

## GVTAFLMUSABEYPVSSGARDLQRKGLRLSRKCYAGLSGGGAVALFIRSLCAQPWG

## RPCASSCLQCFLWTEEGPACRPAPVPTFARLRVFNRRKRLLYKRHNLAQVR"

## ORIGIN

Query Match 91.1%; Score 884.4; DB 9; Length 4041;

Best Local Similarity 99.2%; Pred No. 1.4e-130;

Matches 901; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

QY 1 CAGCGGGCGGCGCTGGCGCCAGATCATCCCTACCAACCCCAAGGACAGACAGT 60

Db 3140 CAGCGGGCGGCGCTGGCGCCAGATCATCCCTACCAACCCCAAGGACAGACAGT 3199

QY 61 GCTGCGCGAATACAGAGGCGCTCAGGGCGCTCGCGCACCCCGCACCTGGCCAGTGCACGC 120

```
Db 3200 GCTGGCGGAATACAGAGGCCCTCAAGGGCTCGGCCACCGCACCTGGCCACGCTGCACGC 3259
QY 121 AGCTTACCTCAGCCCGCCGACACCTGGTGTCTCATCTTGAGGTGTGTCTCTGGGCCGAGCT 180
Db 3260 AGCTTACCTCAGCCCGCCGACACCTGGTGTCTCATCTTGAGGTGTGTCTCTGGGCCGAGCT 3319
QY 191 GCTCCCTCTGCTGGCCGAGAGGGCTCTCTACTCAGATCCAGGTGAAGACTACCTGTG 240
Db 3320 GCTCCCTCTGCTGGCCGAGAGGGCTCTCTACTCAGATCCAGGTGAAGACTACCTGTG 3379
QY 241 GCAGATGTTAGTGTCCACCCAGTACCTGSCAACACAGCACATCTGCACCTGGACCTGAG 300
Db 3380 GCAGATGTTAGTGTCCACCCAGTACCTGSCAACACAGCACATCTGCACCTGGACCTGAG 3439
QY 301 GTCGAGAACATGATCATCCGGAATACAACTGCTCAAGTCTGAGCTGGCAATGTC 360
Db 3440 GTCGAGAACATGATCATCCGGAATACAACTGCTCAAGTCTGAGCTGGCAATGTC 3499
QY 361 ACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAGTTCAGAGCTACCTAGAGAC 420
Db 3500 ACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAGTTCAGAGCTACCTAGAGAC 3559
QY 421 CATGCTCCAGAGCTCCTGGAGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 480
Db 3560 CATG-----GAGCTCCTGGAGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 3613
QY 481 CGGTGTGACACCTTCATCTGAGGCCCGAGTACCCGTTGAGCAGCAGAGGTGACG 540
Db 3614 CGGTGTGACACCTTCATCTGAGGCCCGAGTACCCGTTGAGCAGCAGAGGTGACG 3673
QY 541 CGACTGTGAGAGAGACTGCGCAAGGGCTGGTCCGGCTGAGCCGCTGCTACGCGGGCT 600
Db 3674 CGACTGTGAGAGAGACTGCGCAAGGGCTGGTCCGGCTGAGCCGCTGCTACGCGGGCT 3733
QY 601 GTCGCGGGCGCGTGGCTTCTTCTGCGAGCAGTCTGTGCGCCCGAGCCCTGGGGCGGGC 660
Db 3734 GTCGCGGGCGCGTGGCTTCTTCTGCGAGCAGTCTGTGCGCCCGAGCCCTGGGGCGGGC 3793
QY 661 CTGCGCGTCCAGCTGCTGCGAGTCCGCTGAGTCCGCTGAGCCGCTGCTACGCGGGCT 720
Db 3794 CTGCGCGTCCAGCTGCTGCGAGTCCGCTGAGTCCGCTGAGTCCGCTGAGTCCGCTGCTGCG 3853
QY 721 GCCCGCGCGGTGACCTTCCCTACCGCGGGCTGCGGGCTTCTGCGGCAATCGCGAGAA 780
Db 3854 GCCCGCGCGGTGACCTTCCCTACCGCGGGCTGCGGGCTTCTGCGGCAATCGCGAGAA 3913
QY 781 GAGACGCGGTGCTGTACAGAGGACACACTGCGCCAGGTGCGCTGAGGCTGCGCCCG 840
Db 3914 GAGACGCGGTGCTGTACAGAGGACACACTGCGCCAGGTGCGCTGAGGCTGCGCCCG 3973
QY 841 GCCACACCTTGGTCTCCCGCTGGGGTCCGCTGACAGCGCGCCAAATAAAAAAGCCCGAGC 900
Db 3974 GCCACACCTTGGTCTCCCGCTGGGGTCCGCTGACAGCGCGCCAAATAAAAAAGCCAGC 4033
QY 901 CGGCGGAG 908
Db 4034 CGGCGGAG 4041

RESULT 11
LOCUS BC046431 1183 bp mRNA linear ROD 16-SEP-2003
DEFINITION Mus musculus cDNA sequence BC046431, mRNA (cDNA clone IMAGE:4006702), partial cds.
ACCESSION BC046431
VERSION BC046431.1 GI:28386252
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1183)
```

## AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, E., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Schetzl, J.B., Brownstein, M.J., Udwin, L.B., Toshiyuki, S., Carrincci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.C., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kesteven, M., Maman, A., Rodriguez, S., Bouffard, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bontar, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

## TITLE

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22398257

12477932

2 (bases 1 to 1183)

Strausberg, R.

Direct Submission

## AUTHORS

Submitted (03-FEB-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: [ang@bcm.tmc.edu](mailto:ang@bcm.tmc.edu)

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louleseg, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAX Plate: 92 Row: d Column: 18.

## FEATURES

source

1. 1183  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CZECH II"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4006702"  
/tissue\_type="Mammary tumor metastasized to lung."  
MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR enhancer."

/clone\_lib="NCI CGAP\_Lu30"  
/lab\_host="DH10B"  
/note="vector: PCMV-SPORT6"  
<1. 1183  
/gene="BC046431"  
/db\_xref="LocusID:216790"  
/db\_xref="MGI:2664530"  
<1. 1027  
/gene="BC046431"  
/codon\_start=2  
/product="BC046431 protein"  
/protein\_id="AAH46431.1"  
/db\_xref="GI:28386253"  
/db\_xref="LocusID:216790"  
/db\_xref="MGI:2664530"

## gene

BC046431 1183 bp mRNA linear ROD 16-SEP-2003  
Mus musculus cDNA sequence BC046431, mRNA (cDNA clone IMAGE:4006702), partial cds.

## CDS

BC046431.1 GI:28386252

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 1183)



```
/translation="VSKAGMGYPSSPEQVLGGPNHLASEEBSRRPAQLLPSTKT
FAFOMQRRGRFSVRCREKASGRALAAKIVYQEDDKTAVLRVEALKRLHHPHLA
QLHAAYLSPRLHVLIELCSPELLSLAESYSSESDVDKLDWQMSATVYLHAQHI
LHLDLRSENMMVTEYNLLKVIDLGNASDOEKVPAPENFKDYLETMAFELLEGQAV
PQIDIAWIGTATFIMLSGEYFESSEGTDLQKRLKGLRLSCRYAGLSGGAFAFLQS
SLCAQPMGRPCASTCLQCGWLTBEGPTGSRPTVPTPTVLRAPFVREKRALLYKK
HNLAQVR"
misc_feature
134..892
/gene="BC046431"
/note="S-Trk; Region: Serine/Threonine protein kinases,
catalytic domain"
/db_xref="CDD:smart00220"

ORIGIN
Query Match 65.1%; Score 631.8; DB 10; Length 1183;
Best Local Similarity 79.2%; Pred. No. 1.7e-90;
Matches 763; Conservative 0; Mismatches 197; Indels 3; Gaps 1;

Qy 2 AGCGGGCGGGCGCTGGCGCCAGATCATCCCTTACCACCCCAAGGACAGACAGCAGTG 61
Db 200 AGTGGCGGGCCCTAGCTGTCTAGATCTTCCCTTACCACCTGAGGACAGACAGCTGTA 259
Qy 62 CTGCGGGAATACGAGGCGCTCAAGGCGCTGGCCACCCGACCTGGCCAGCTGCAGCA 121
Db 260 CTAAGAGAATACGAGGCGCTTAAAGAGCTGCACACCCACATCTGGCCCACTCCATGCC 319
Qy 122 GCCTACCTCAGCCCGCCGACCTGGTCTCATCTTGGAGCTGTGCTCTGGCCCGAGCTG 181
Db 320 GCCTACCTCAGTCCCGGACCTGGTCTCATCTGGAGCTGTGCTCTGGCCCTGAGCTG 379
Qy 182 CTCCTCTGCTGGCCGAGAGGCGCTCTACTCAGATTCGAGGTGAAGACTACCTGTGG 241
Db 380 CTACCTCTCTTGGCGGAGAGGAACTCTACTCAGAGTCTGATGTGAAGGACTACCTGTGG 439
Qy 242 CAGATGTTGAGTGCACCCAGTACTCTGCACAAACAGACACATCTGCACTGGACCTGAGG 301
Db 440 CAGATGTTGAGTGCACCCAGTACTCTGCATGCCCAACACATCTGCACCTGGACCTGAGG 499
Qy 302 TCCGAGAACATGATCATCAGCAATACAACTGCTCAAGCTGTGGACCTGGGCAATGCA 361
Db 500 TCCGAGAACATGATGCTCACTGAGTACAACTGCTTAAAGTTATAGACCTGGGAAATGCC 559
Qy 362 CAGAGCTTCAGCAGGAGAGGTCTGCTCCCTCAGACAGTTCAGGACTACCTAGAGACC 421
Db 560 CAGAGTCTCGACCAAGAGAGGTCCAGCCCTGAGAACTTCAAGAACTACCTGGAGACC 619
Qy 422 ATGGCTCCAGAGCTCTCGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCGATC 481
Db 620 ATGGCTCCAGAACTCTCGAAGGTCAAGGGCTGTTCCACAGACAGACATTTGGGCTATT 679
Qy 482 GGTGTACAGCTTCATCTGCTGAGCGCGAGTACCCGCTGAGCAGCAGGCTGCACGC 541
Db 680 GGTGTACAGCTTCATCTGCTGAGTGGCGAGTACCCAGAGACAGCAGGAGGACTCGT 739
Qy 542 GACCTGCAGAGAGACTGCGCAAGGGCTGTGCTCGGCTGAGCGGCTGTCTACCGGGGGCTG 601
Db 740 GACCTGCAGAAAGCTTGGCAAGGGACTCATTCGGTTGAGTGCCTGCTATGCAGGATTA 799
Qy 602 TCCGGGGCGCGCTGCTGCTGCGGAGCATCTGTGCGCCCGACGCGCTGGGCGCGGCC 661
Db 800 TCAGGAGAGCGGTAGCTTCTGCTGAGAGTTTCAATATGTCTCAACCTGGGCGCGCCG 859
Qy 662 TGCCTCTCAGCTCCCTGCTGAGTGCCTGCTTACAGAGAGGCGCGGCTGTTCGCGG 721
Db 860 TGCCTCTCAGCTCTGCAATGCGGTGCTGACAGAGAGGCGCGGCTCCGCTCCCG 919
Qy 722 CCGCGCGCGGTGACCTTCCCTTACCGCGCGGTGCGCTCTTCGTGCGCAATTCGCGAAG 781
Db 920 CCCACGCGCGTAACTTCCCGACCGCTGCGGCTGCGGCTGCGGCTTGTGCGCGAGGAG 979
Qy 782 AGAGCGCGCTGTGACAGAGGACAACTGCGCCCGAGGTGCGGTGAGGCTGCGCCCGG 841
Db 980 CGGCGTGGCTCTCTTACAGAGACATTAACCTGCGCCCGAGGTGCGGTGAGG---TCCAGCT 1036
```

```
QY 842 CCACACCTTGGTCTCCCGCTGGGGTCTGCAGACGGCCCAATAAAACGCCGCC 901
Db 1037 CTACAGACAAGATGTCATCGGCAGACCCCTTAATAAGATCAAAAC 1096
QY 902 GGGCGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 961
Db 1097 GGGCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1156
QY 962 AAA 964
Db 1157 AAA 1159

RESULT 12
AL670729
LOCUS
DEFINITION Human DNA sequence from clone RP11-245P10 on chromosome 1, complete
sequence.
ACCESSION AL670729
VERSION AL670729.19 GI:22316158
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 62164)
AUTHORS Van Hellmond,Z.
TITLE Direct Submision
JOURNAL Submitted (18-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
clonerequest@sanger.ac.uk
humquery@sanger.ac.uk
On Aug 19, 2002 this sequence version replaced GI:21727388.
COMMENT
-----
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information on the WormPep
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-245P10 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
Location/Qualifiers
1..62164
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-245P10"
/clone_lib="RPCI-11.1"
```



## ORIGIN

Query Match 27.0%; Score 262; DB 9; Length 62164;  
Best Local Similarity 98.1%; Pred. No. 3.8e-32;  
Matches 265; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 641 GCCCAGCCCTGGGCGCCCTGGCGTCCAGCTCCCTGAGTCCCGTGGCTTAACAGAG 700  
DB 31970 GTCCGTCCAGAGGGCGCCCTGGCGTCCAGCTCCCTGAGTCCCGTGGCTTAACAGAG 32029

QY 701 GAGGCGCCGGCTGTTCGCGCGCGCGCGCCGCTGACCTTCCCTACCGCGGCTCGCGGTC 760  
DB 32030 GAGGCGCCGGCTGTTCGCGCGCGCGCGCCGCTGACCTTCCCTACCGCGGCTCGCGGTC 32089

QY 761 TTCGTGCCCAATCGCGAAGAGAGACGCGCGTGTGTACAGAGGCAACCTGGCCAG 820  
DB 32090 TTCGTGCCCAATCGCGAAGAGAGACGCGCGTGTGTACAGAGGCAACCTGGCCAG 32149

QY 821 GTGCGCTGAGGTGCGCCCGCCACACACCTTGCTCCCGCTGGGCTGCTGCACAGC 880  
DB 32150 GTGCGCTGAGGTGCGCCCGCCACACCTTGCTCCCGCTGGGCTGCTGCACAGC 32209

QY 881 CGCCAATAAAAAACCGCCAGCGCGCGAGAA 910  
DB 32210 CGCCAATAAAAAACCGCCAGCGCGCGAGAA 32239

## RESULT 13

AC026657/c  
LOCUS Homo sapiens chromosome 1 clone RP11-245P10, WORKING DRAFT  
DEFINITION AC026657 164766 bp DNA linear HTG 01-SEP-2000  
SEQUENCE, 31 unordered pieces.  
ACCESSION AC026657  
VERSION AC026657.4 GI:9958202  
KEYWORDS HTG; HIGS\_PHASE1; HIGS\_DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 164766)  
Waterston,R.H.  
The sequence of Homo sapiens clone  
2 (bases 1 to 164766)  
Waterston,R.H.  
Direct Submission  
Submitted (22-MAR-2000) Genome Sequencing Center, Washington  
University School of Medicine, 444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Sep 1, 2000 this sequence version replaced gi:7637349.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----  
Center project name: H.NH024P10

----- Summary Statistics -----  
Sequencing vector: M13, 100%

Sequencing vector: plasmid, 0%

Chemistry: Dye-terminator Big Dye, 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 139884 bases at least Q40

Consensus quality: 147686 bases at least Q30

Consensus quality: 151469 bases at least Q20

Insert size: 169000; agarose-ff

Insert size: 161074; sum-of-contigs

Quality coverage: 3.60 in Q20 bases; agarose-ff

Quality coverage: 3.92 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 31 contigs. The true order of the pieces

\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1124: contig of 1124 bp in length  
1125: gap of unknown length  
1224: contig of 1808 bp in length  
1225: gap of unknown length  
3032: contig of 1361 bp in length  
3033: gap of unknown length  
3132: contig of 1361 bp in length  
3133: gap of unknown length  
4494: contig of 1267 bp in length  
4494: gap of unknown length  
4594: contig of 1267 bp in length  
4594: gap of unknown length  
5861: contig of 1710 bp in length  
5861: gap of unknown length  
7671: contig of 1899 bp in length  
7671: gap of unknown length  
9670: contig of 1946 bp in length  
9670: gap of unknown length  
11715: contig of 1946 bp in length  
11715: gap of unknown length  
14243: contig of 2428 bp in length  
14243: gap of unknown length  
14344: contig of 2544 bp in length  
14344: gap of unknown length  
16888: contig of 2544 bp in length  
16888: gap of unknown length  
19247: contig of 2260 bp in length  
19247: gap of unknown length  
19348: contig of 2028 bp in length  
19348: gap of unknown length  
21376: contig of 3550 bp in length  
21376: gap of unknown length  
25025: contig of 3550 bp in length  
25025: gap of unknown length  
25126: contig of 2926 bp in length  
25126: gap of unknown length  
28052: contig of 3903 bp in length  
28052: gap of unknown length  
28152: contig of 3903 bp in length  
28152: gap of unknown length  
32055: contig of 3562 bp in length  
32055: gap of unknown length  
35717: contig of 3562 bp in length  
35717: gap of unknown length  
40891: contig of 5075 bp in length  
40891: gap of unknown length  
40992: contig of 3036 bp in length  
40992: gap of unknown length  
44028: contig of 4772 bp in length  
44028: gap of unknown length  
48999: contig of 5256 bp in length  
48999: gap of unknown length  
54255: contig of 5256 bp in length  
54255: gap of unknown length  
54355: contig of 4639 bp in length  
54355: gap of unknown length  
58994: contig of 4639 bp in length  
58994: gap of unknown length  
59095: contig of 5296 bp in length  
59095: gap of unknown length  
64390: contig of 6375 bp in length  
64390: gap of unknown length  
70865: contig of 7702 bp in length  
70865: gap of unknown length  
78667: contig of 7702 bp in length  
78667: gap of unknown length  
87207: contig of 8440 bp in length  
87207: gap of unknown length  
87307: contig of 10551 bp in length  
87307: gap of unknown length  
97958: contig of 11562 bp in length  
97958: gap of unknown length  
109521: contig of 11562 bp in length  
109521: gap of unknown length  
120928: contig of 11288 bp in length  
120928: gap of unknown length  
121008: contig of 20469 bp in length  
121008: gap of unknown length  
141577: contig of 21095 bp in length  
141577: gap of unknown length  
162672: contig of 21095 bp in length  
162672: gap of unknown length  
162673: contig of 1202 bp in length  
162673: gap of unknown length  
163975: contig of 692 bp in length  
163975: gap of unknown length  
164075: contig of 692 bp in length  
164075: gap of unknown length

## FEATURES

source

1. 164766  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="1"

```
misc_feature      /clone="RP11-245P10"
1. .1124
/note="assembly_name:Contig14"
misc_feature      1225. .3032
/note="assembly_name:Contig21"
misc_feature      3133. .4493
/note="assembly_name:Contig23"
misc_feature      4594. .5860
/note="assembly_name:Contig24"
misc_feature      5961. .7670
/note="assembly_name:Contig25"
misc_feature      7771. .9669
/note="assembly_name:Contig26"
misc_feature      9770. .11715
/note="assembly_name:Contig27"
misc_feature      11816. .14243
/note="assembly_name:Contig28"
misc_feature      14344. .16887
/note="assembly_name:Contig29"
misc_feature      16988. .19247
/note="assembly_name:Contig30"
misc_feature      19348. .21375
/note="assembly_name:Contig31"
misc_feature      21476. .25025
/note="assembly_name:Contig32"
misc_feature      25126. .28051
/note="assembly_name:Contig33"
misc_feature      28152. .32054
/note="assembly_name:Contig34"
misc_feature      32155. .35716
/note="assembly_name:Contig35"
misc_feature      35817. .40891
/note="assembly_name:Contig36"
misc_feature      40992. .44027
/note="assembly_name:Contig37"
misc_feature      44128. .48899
/note="assembly_name:Contig38"
misc_feature      49000. .54255
/note="assembly_name:Contig39"
misc_feature      54356. .58994
/note="assembly_name:Contig40"
misc_feature      59095. .64390
/note="assembly_name:Contig41"
misc_feature      64491. .70865
/note="assembly_name:Contig42"
misc_feature      70966. .78667
/note="assembly_name:Contig43"
misc_feature      78768. .87207
/note="assembly_name:Contig44"
misc_feature      87308. .97858
/note="assembly_name:Contig45"
misc_feature      97959. .109520
/note="assembly_name:Contig46"
misc_feature      109621. .120908
/note="assembly_name:Contig47"
misc_feature      121009. .141477
/note="assembly_name:Contig48"
misc_feature      141578. .162672
/note="assembly_name:Contig49"
misc_feature      162773. .163974
/note="assembly_name:Contig99"
misc_feature      164075. .164766
/note="assembly_name:Contig7"

ORIGIN
Query Match      23.6%; Score 229.4; DB 2; Length 164766;
Best Local Similarity 97.5%; Pred. No. 4.9e-27;
Matches 233; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY      191 CTGGCCGAGAGGGCTCTCTACTCAGATCCGAGTGAAGGACTACCTGTGGCAGATGTTG 250
Db      65348 CTGCCCCCGAGGGCTCTCTACTCAGATCCGAGTGAAGGACTACCTGTGGCAGATGTTG 65289

QY      251 AGTGCCACCCAGTACCTGCACACAGCACATCTCGACCTGGACCTGAGGTCCGAGAAC 310
Db      65288 AGTGCCACCCAGTACCTGCACACAGCACATCTCGACCTGGACCTGAGGTCCGAGAAC 65229

QY      311 ATGATCATCACCGAATACACCTGCTCAAGTCTGTGGACCTGGCCATGACAGAGCTC 370
Db      65228 ATGATCATCACCGAATACACCTGCTCAAGTCTGTGGACCTGGCCATGACAGAGCTC 65169

QY      371 AGCCAGGAGAGAGTGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGACCATGGCTCC 429
Db      65168 AGCCAGGAGAGAGTGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGACCATGGGTGC 65110

AC023889      174612 bp      DNA      linear      HTG 07-JUL-2000
Homo sapiens chromosome 1 clone RP11-661B12, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC023889      GI:8969253
HTG: HTGS_PHASE1; HTGS_DRAFT.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
1 (bases 1 to 174612)
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 174612)
Waterston,R.H.
Direct Submission
Submitted (18-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jul 7, 2000 this sequence version replaced gi:8748947.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0661B12
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 163945 bases at least Q40
Consensus quality: 167601 bases at least Q30
Consensus quality: 169687 bases at least Q20
Insert size: 148000; agarose-fp
Insert size: 173412; sum-of-contigs
Quality coverage: 5.67 in Q20 bases; agarose-fp
Quality coverage: 5.55 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2478: contig of 2478 bp in length
* 2479 2578: gap of unknown length
* 2579 6948: contig of 4370 bp in length
* 7048 7049: gap of unknown length
* 9373 9473: contig of 2325 bp in length
* 9474 13303: gap of unknown length
* 13303 13403: contig of 3830 bp in length
* 13403 13403: gap of unknown length
```

\* 13404 18270: contig of 4867 bp in length  
\* 18271 18370: gap of unknown length  
\* 18371 22589: contig of 7219 bp in length  
\* 25690 25689: gap of unknown length  
\* 25690 34815: contig of 9126 bp in length  
\* 34816 34915: gap of unknown length  
\* 34916 45742: contig of 10827 bp in length  
\* 45743 45842: gap of unknown length  
\* 45843 62987: contig of 17145 bp in length  
\* 62988 63087: gap of unknown length  
\* 63088 87584: contig of 24497 bp in length  
\* 87585 87684: gap of unknown length  
\* 87685 114765: contig of 27081 bp in length  
\* 114766 114865: gap of unknown length  
\* 114866 142334: contig of 27469 bp in length  
\* 142335 142434: gap of unknown length  
\* 142435 174612: contig of 32178 bp in length.

## FEATURES

Location/Qualifiers

source

1. 174612  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/clone="RP11-661B12"  
misc\_feature  
1. 2478  
/note="assembly\_name:Contig4"  
misc\_feature  
2575. 6948  
/note="assembly\_name:Contig5"  
misc\_feature  
7049. 9373  
/note="assembly\_name:Contig6"  
misc\_feature  
9474. 13303  
/note="assembly\_name:Contig7"  
misc\_feature  
13404. 18270  
/note="assembly\_name:Contig8"  
misc\_feature  
18371. 25589  
/note="assembly\_name:Contig9"  
misc\_feature  
25690. 34815  
/note="assembly\_name:Contig10"  
misc\_feature  
34916. 45742  
/note="assembly\_name:Contig11"  
misc\_feature  
45843. 62987  
/note="assembly\_name:Contig12"  
clone\_end:SP6  
vector\_side:right  
63088. 87584  
/note="assembly\_name:Contig13"  
misc\_feature  
87685. 114765  
/note="assembly\_name:Contig14"  
misc\_feature  
114866. 142334  
/note="assembly\_name:Contig15"  
misc\_feature  
142435. 174612  
/note="assembly\_name:Contig16"

## ORIGIN

Query Match 23.6%; Score 229.4; DB 2; Length 174612;  
Best Local Similarity 97.5%; Pred. No. 4.8e-27;  
Matches 233; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 191 CTGGCCGAGAGGGCTCTACTCAGAAATCCGAGGTGAAGGACTACCTGTGGCAGATGTTG 250  
Db 19339 CTGCCCCCAGGGGCTCTCTACTCAGAAATCCGAGGTGAAGGACTACCTGTGGCAGATGTTG 19280  
QY 251 AGTGCCACCCAGTACTCGACAAACAGACATCTCTGACCTGGACCTGAGGTCGAGAAC 310  
Db 19279 AGTGCCACCCAGTACTCGACAAACAGACATCTCTGACCTGGACCTGAGGTCGAGAAC 19220  
QY 311 ATGATCATCCGATACAACTGCTCAAGTCTGAGCTGGGCAATGACAGGCTC 370  
Db 19219 ATGATCATCCGATACAACTGCTCAAGTCTGAGCTGGGCAATGACAGGCTC 19160  
QY 371 AGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGACCATGGCTCC 429  
Db 19159 AGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGACCATGGGTGC 19101

## RESULT 15

AX309418  
LOCUS AX309418 225 bp DNA linear PAT 14-DEC-2001  
DEFINITION Sequence 2403 from Patent WO0190366.  
ACCESSION AX309418  
VERSION AX309418.1 GI:17895939  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1  
AUTHORS Leach,M.D. and Shimkets,R.A.  
TITLE Human polynucleotides and polypeptides encoded thereby  
JOURNAL Patent: WO 0190366-A 2403 29-NOV-2001;  
Curagen Corporation (US)  
FEATURES  
Location/Qualifiers  
source  
1. .225  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 23.0%; Score 223.4; DB 6; Length 225;  
Best Local Similarity 99.6%; Pred. No. 1e-25;  
Matches 224; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 314 ATCATCACCGAATACAACTGCTCAAGTCTGGACCTGGGCAATGCACAGACCTTCAGC 373  
Db 1 ATCATCACCGAATACAACTGCTCAAGTCTGGACCTGGGCAATGCACAGACCTTCAGC 60  
QY 374 CAGGAGAAGTGCTGCCCTCAGACAAAGTTCAGGACTACCTAGAGACCATGGCTCCAGAG 433  
Db 61 CAGGAGAAGTGCTGCCCTCAGACAAAGTTCAGGACTACCTAGAGACCATGGCTCCAGAG 120  
QY 434 CTCCTGGAGGGCCAGGGGCTGTTCACACAGACAGACATCTGGGCCATCGGTGTGACAGCC 493  
Db 121 CTCCTGGAGGGCCAGGGGCTGTTCACACAGACAGACATCTGGGCCATCGGTGTGACAGCC 180  
QY 494 TTCATCATGCTAGCCCGAGTACCCGGTGAGCAGCGAGGGTGCA 538  
Db 181 TTCATCATGCTAGCCCGAGTACCCGGTGAGCAGCGAGGGTGCA 225

Search completed: September 19, 2004, 19:26:51

Job time : 3971.31 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 12:53:28 / Search time 400.361 Seconds  
(without alignments)  
10303.209 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_23150\_24120  
Perfect score: 971  
Sequence: 1 cagcggcgggcgctggccg.....ggcgcgtataaaagtctaga 971

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_29Jan04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002s:\*
- 7: Geneseqn2003as:\*
- 8: Geneseqn2003bs:\*
- 9: Geneseqn2003cs:\*
- 10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	971	100.0	8106	7 ABX11641	Abx11641 Human ser
2	971	100.0	24120	7 ABX11642	Abx11642 Human ser
3	946.4	97.5	4175	5 AAC62285	Aac62285 cDNA enco
4	926.8	95.4	5207	6 AAL43909	Aal43909 Human kin
5	926.8	95.4	5207	6 AAL43908	Aal43908 Human kin
6	911.4	93.9	3225	6 AAD26467	Aad26467 Human kin
7	911.4	93.9	5454	6 AAD38865	Aad38865 Human kin
8	905.4	93.3	4936	4 AAH46904	Aah46904 cDNA enco
9	905.4	93.2	5007	5 AAC62286	Aac62286 cDNA enco
10	905.4	93.2	7928	5 AAC62287	Aac62287 cDNA enco
11	223.4	23.0	225	6 ABN76255	Abn76255 Human kin
12	206	21.2	7710	4 AAF44862	Aaf44862 Novel pro
13	204.4	21.1	7789	6 AAD30565	Aad30565 Human kin
14	204.4	21.1	9807	9 AAD79958	Aad79958 Human kin
15	204.4	21.1	10662	9 ADC99127	Adc99127 Human KPP
16	202.8	20.9	9698	9 ADE47675	Ade47675 Human NOV
17	202.8	20.9	10122	9 ADE47673	Ade47673 Human NOV
18	135.6	14.0	9930	9 ADE47671	Ade47671 Human NOV
19	126.8	13.1	182	3 AAC25654	Aac25654 Human sec
20	113.8	11.7	2079	4 AAK94258	Aak94258 Human ful
21	113.8	11.7	2224	4 AAH16158	Aah16158 Human cDN
22	113.8	11.7	2224	5 AAH78068	Aah78068 Nucleotid
23	113.8	11.7	2226	9 ADD29770	Add29770 Human tum

24	112.2	11.6	2105	6 ABV72291	Abv72291 Nucleotid
25	112.2	11.6	2132	2 AAX34656	Aax34656 Human ZIP
26	112.2	11.6	2132	7 ACA90235	Ac90235 Deatch as
27	94.4	9.7	1514	9 ADB58463	Adb58463 Toxicity-
28	94.4	9.7	1514	9 ADB53052	Adb53052 Primary r
29	94.2	9.7	1429	2 AAX34657	Aax34657 Murine ZI
30	90.4	9.3	757	4 AAX93262	Aax93262 Human cDN
31	90.4	9.3	757	4 AAX91856	Aax91856 Human cDN
32	89.2	9.2	1791	6 AAD36211	Aad36211 Human car
33	89.2	9.2	1791	6 AAD36212	Aad36212 Human car
34	89.2	9.2	1791	6 AAD36209	Aad36209 Human car
35	89.2	9.2	1791	6 AAD36040	Aad36040 Human car
36	89.2	9.2	1791	6 AAD36210	Aad36210 Human car
37	89.2	9.2	1836	4 AAS06721	Aas06721 Polynucle
38	89.2	9.2	2046	5 AAH78263	Aah78263 Nucleotid
39	89	9.2	860	9 ADE47677	Ade47677 Human NOV
40	87.6	9.0	1839	6 AAD26573	Aad26573 Human POL
41	84.6	8.7	1788	6 AAD26570	Aad26570 Human POL
42	84.6	8.7	1971	6 AAD26571	Aad26571 Human POL
43	84.6	8.7	2558	6 AAD26572	Aad26572 Human POL
44	83.2	8.6	1032	4 AAS06711	Aas06711 Polynucle
45	83.2	8.6	1216	6 ABZ11323	Abz11323 Human pol

## ALIGNMENTS

RESULT 1  
ABX11641  
ID ABX11641 standard; cDNA; 8106 BP.

XX  
AC ABX11641;

XX  
DT 09-MAY-2003 (first entry)

XX  
DE Human serine/threonine or protein kinase 59079, cDNA.

XX  
KW Human; ss; gene; serine/threonine kinase; protein kinase; 59079;  
KW cardiovascular disease; heart failure; myocardial infarction;  
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;  
KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;  
KW haemolytic anaemia; cellular proliferative disorder; cancer;  
KW protein kinase disorder; autoimmune disorder; diabetes mellitus;  
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;  
KW multiple sclerosis.

XX  
OS Homo sapiens.

XX  
FH Key Location/Qualifiers  
FT 5'UTR 1..71  
FT CDS /\*tag= a  
FT 72..7964 /\*tag= b  
FT /\*product= "Kinase 59079"  
FT 3'UTR /\*note= "This CDS is specifically claimed in claim 2"  
FT 7965..8106 /\*tag= c

XX  
US2002168742-A1.

XX  
PD 14-NOV-2002.

XX  
PF 15-FEB-2002; 2002US-00077130.

XX  
PR 15-FEB-2001; 2001US-0269201P.

XX  
PA (WILL-) MILLENNIUM PHARM INC.

XX  
PI Kapeller-Libermann R, Acton SL;

XX  
DR WPI; 2003-298729/29.

XX  
DR P-PSDB; ABG76186.



PF 15-FEB-2002; 2002US-00077130.  
XX  
PR 15-FEB-2001; 2001US-0269201P.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Kapeller-Libermann R, Acton SL;  
XX  
DR WPI; 2003-298729/29.  
DR P-PSDB; ABG76187.  
XX  
PT Novel isolated human protein kinase, designated 59079 or 12599  
PT polypeptide, useful as diagnostic and therapeutic agents for preventing  
PT cardiovascular diseases, proliferative disorders, and protein kinase  
PT disorders.  
XX  
PS Claim 2; Page 58-84; 119pp; English.  
XX  
CC The invention relates to an isolated human serine/threonine or protein  
CC kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule  
CC comprising at least 85% identity to the nucleic acids appearing as  
CC ABX11641 and ABX11642 or their complement, a naturally occurring variant  
CC of the kinases or their fragments. Also included are a non-human host  
CC cell containing the nucleic acids, an antibody specific for the proteins,  
CC identifying a compound which binds to the kinase (by contacting the  
CC kinase or a cell expressing the kinase with a test compound and  
CC determining whether the kinase binds to the test compound) and modulating  
CC the activity of kinase using the identified compound. The kinases and  
CC their encoding nucleic acids are useful as diagnostic and therapeutic  
CC agents for preventing a disease or condition associated with an aberrant  
CC or unwanted 59079 or 12599 activity in a subject, including  
CC cardiovascular diseases such as heart failure, and myocardial infarction;  
CC disorders involving blood vessels such as atherosclerosis, and Kaposi's  
CC sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia,  
CC Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders  
CC such as cancer; and protein kinase disorders such as autoimmune  
CC rheumatoid arthritis, and multiple sclerosis (many examples of diseases  
CC and disorders are included in the specification). The kinases, their  
CC encoding nucleic acids and antibodies are useful in screening assays,  
CC detection assays (e.g. forensic biology), and predictive medicine (e.g.  
CC diagnostic assays), prognostic assays, and monitoring clinical trials and  
CC pharmacogenomics). The kinases and their encoding nucleic acids are  
CC useful as query sequences to perform a search against public databases to  
CC identify other family members or related sequences. The present sequence  
XX encodes the kinase 12599  
SQ Sequence 24120 BP; 4719 A; 7265 C; 8278 G; 3858 T; 0 U; 0 Other;

Query Match 100.0%; Score 971; DB 7; Length 24120;  
Best Local Similarity 100.0%; Pred. No. 1.4e-171;  
Matches 971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCGGGGGGGCTGGCGGCCAGATCATCCCTTACCACCCCAAGGACAGACGAGT 60  
DB 23150 CAGCGGGGGGGCTGGCGGCCAGATCATCCCTTACCACCCCAAGGACAGACGAGT 23209

QY 61 GCTGCGGAATACAGAGGCGCTCAAGGGCTGGCGCCACCCGACCTGGCCCAAGTGCACGC 120  
DB 23210 GCTGCGGAATACAGAGGCGCTCAAGGGCTGGCGCCACCCGACCTGGCCCAAGTGCACGC 23269

QY 121 AGCTTACTCAGCCCGCCGACCTGGCTCTATCTTGAGTGTGCTTGGCGCCGAGCT 180  
DB 23270 AGCTTACTCAGCCCGCCGACCTGGCTCTATCTTGAGTGTGCTTGGCGCCGAGCT 23329

QY 181 GTCTCCCTGCTGGCGGAGAGGGCTCTCTACTCAGATCCGAGGTGAAGGACTACCTGTG 240  
DB 23330 GTCTCCCTGCTGGCGGAGAGGGCTCTCTACTCAGATCCGAGGTGAAGGACTACCTGTG 23389

QY 241 GCAGATGTGAGTGCACCCAGTACTCTGCACCAACAGACATCTCTGCACCTGGACCTGAG 300  
DB 23390 GCAGATGTGAGTGCACCCAGTACTCTGCACCAACAGACATCTCTGCACCTGGACCTGAG 23449

QY 301 GTCCGAGAACATGATCATCACCAGATACAACTGCTCAAGTCTGCTGGACCTGGCAATGC 360  
DB 23450 GTCCGAGAACATGATCATCACCAGATACAACTGCTCAAGTCTGCTGGACCTGGCAATGC 23509

QY 361 ACAGAGCCTCAGCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACTAGAGAC 420  
DB 23510 ACAGAGCCTCAGCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACTAGAGAC 23569

QY 421 CATGGCTCCAGAGCTCTCTGAGAGGCGCAGGGGGCTGTTCACAGACAGACATCTGGGCCAT 480  
DB 23570 CATGGCTCCAGAGCTCTCTGAGAGGCGCAGGGGGCTGTTCACAGACAGACATCTGGGCCAT 23629

QY 481 CGGTGTGACAGCTTCATCATCTGAGCGCCGAGTACCCGGTGTGAGCAGGAGGTGCACG 540  
DB 23630 CGGTGTGACAGCTTCATCATCTGAGCGCCGAGTACCCGGTGTGAGCAGGAGGTGCACG 23689

QY 541 CGACCTGCAGAGAGGACTCCGCAAGGGGCTGTGCCGGTGTGAGCCGCTGTCTACGCGGGCT 600  
DB 23690 CGACCTGCAGAGAGGACTCCGCAAGGGGCTGTGCCGGTGTGAGCCGCTGTCTACGCGGGCT 23749

QY 601 GTCCGGGGGGCGGCTTCTTCCGCGAGCACTCTGTGCGCCGACGCTTGGGGCGCGGCC 660  
DB 23750 GTCCGGGGGGCGGCTTCTTCCGCGAGCACTCTGTGCGCCGACGCTTGGGGCGCGGCC 23809

QY 661 CTGCGCGCTCAGCTGCTGAGTGCCTGCGCGTGGCTTAACAGAGGAGGGCCCGGCTTTCGCG 720  
DB 23810 CTGCGCGCTCAGCTGCTGAGTGCCTGCGCGTGGCTTAACAGAGGAGGGCCCGGCTTTCGCG 23869

QY 721 GCCCGCGCGCTGACCTTCCCTACCGCGCGGCTTGGCGCTTCTGCTGCGCAATCGCGAGAA 780  
DB 23870 GCCCGCGCGCTGACCTTCCCTACCGCGCGGCTTGGCGCTTCTGCTGCGCAATCGCGAGAA 23929

QY 781 GAGACGCGCGCTGCTCTACAAGAGGCAACAACCTGGCCGAGGTGCGCTGAGGGTTCGCCCCG 840  
DB 23930 GAGACGCGCGCTGCTCTACAAGAGGCAACAACCTGGCCGAGGTGCGCTGAGGGTTCGCCCCG 23989

QY 841 GCCACACCTTGTCTCCCGCTGGGGTGGCTGTGAGAGCGCGCCAAATAAAAGCCGCCAGC 900  
DB 23990 GCCACACCTTGTCTCCCGCTGGGGTGGCTGTGAGAGCGCGCCAAATAAAAGCCGCCAGC 24049

QY 901 CGGCGCAGAAATAAGGCGCGCTAA 960  
DB 24050 CGGCGCAGAAATAAGGCGCGCTAA 24109

QY 961 AAAAGTCTAGA 971  
DB 24110 AAAAGTCTAGA 24120

RESULT 3  
AAC62285  
ID AAC62285 standard; cDNA; 4175 BP.  
XX  
AC AAC62285;  
XX  
DT 19-MAR-2001 (first entry)  
XX  
DE cDNA encoding a human signal transduction polypeptide.  
XX  
KW Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;  
KW congestive heart failure; dilated congestive cardiomyopathy;  
KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;  
KW mitral valve disease; aortic valve disease; tricuspid valve disease;  
KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;  
KW atherosclerosis; cardiac tumour; microbial infection; ss.  
XX  
OS Homo sapiens.  
XX  
PH Key Location/Qualifiers  
FT CDS 1..4056  
FT /\*tag= a  
FT /product= "signal transduction polypeptide H19G5"  
XX

PN WO200063381-A1.  
 XX 26-OCT-2000.  
 XX 11-APR-2000; 2000WO-US009488.  
 XX 16-APR-1999; 99US-0129553P.  
 XX (SCIO-) SCTOS INC.  
 PA Zeng W, Stanton L, Kong H;  
 XX WPI; 2001-007013/01.  
 DR P-PSDB; AAB30567.  
 XX Novel h19G5 polypeptides capable of regulating signal transduction and  
 PT exhibiting kinase activity useful for identifying antibodies to treat  
 PT cardiac diseases, and additional mediators of signal transduction.  
 XX Claim 4; Page 57-59; 81pp; English.  
 CC The present sequence encodes a human protein with putative function in  
 CC signal transduction. The polypeptide is designated H19G5. The protein is  
 CC capable of regulating signal transduction and exhibits kinase activity.  
 CC The H19G5 transcript is expressed in the heart. H19G5 polypeptides and  
 CC polynucleotides are useful for preventing or treating a cardiac disease,  
 CC such as congestive heart failure, dilated congestive cardiomyopathy,  
 CC hypertrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve  
 CC disease, aortic valve disease or tricuspid valve disease, angina  
 CC pectoris, myocardial infarction, cardiac arrhythmia, pulmonary, arterial  
 CC or renovascular hypertension, arteriosclerosis, atherosclerosis and  
 CC cardiac tumours in humans. The polypeptide is also useful for detecting  
 CC the expression of a protein capable of regulating signal transduction or  
 CC the expression of a protein capable of acting as a donor or acceptor  
 CC molecule of a phosphate group. The monoclonal antibodies can be used as  
 CC probes for detecting discrete antigens expressed by tissue or cell  
 CC samples, and therefore used in humans for localization and monitoring of  
 CC microbial infection  
 XX  
 SQ Sequence 4175 BP; 778 A; 1459 C; 1283 G; 655 T; 0 U; 0 Other;  
 Query Match 97.5%; Score 946.4; DB 5; Length 4175;  
 Best Local Similarity 99.9%; Pred. No. 4.1e-167;  
 Matches 947; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGGGCGGCGTGGCCGCAAGATCATCCCTACACCCCAAGCAAGACAGCAGT 60  
 DB 3228 CAGCGGGCGGCGTGGCCGCAAGATCATCCCTACACCCCAAGCAAGACAGCAGT 3287  
 QY 61 GCTGGCGAATACAGAGCCCTCAAGGCGCTGGCCACCCGACCTGGCCCGGACAGC 120  
 DB 3288 GCTGGCGAATACAGAGCCCTCAAGGCGCTGGCCACCCGACCTGGCCCGGACAGC 3347  
 QY 121 AGCCTACTCAGCCCCCGGACCTGGTGTCTCATCTTGGAGCTGTGCTCTGGCCCGGAGCT 180  
 DB 3348 AGCCTACTCAGCCCCCGGACCTGGTGTCTCATCTTGGAGCTGTGCTCTGGCCCGGAGCT 3407  
 QY 181 GCTCCCTGCTGGCCGAGAGGGCTCTCTACTAGAAATCCGAGTGAAGACTACCTGTG 240  
 DB 3408 GCTCCCTGCTGGCCGAGAGGGCTCTCTACTAGAAATCCGAGTGAAGACTACCTGTG 3467  
 QY 241 GCAGATGTGAGTGCCACCCAGTAGTACCTGCACCAACAGCAGCATCTGCACCTGGACCTGAG 300  
 DB 3468 GCAGATGTGAGTGCCACCCAGTAGTACCTGCACCAACAGCAGCATCTGCACCTGGACCTGAG 3527  
 QY 301 GTCCGAGAAATGATCATACCGAATACACCTGCTCAAGGTGTGACCTGGGCAATGC 360  
 DB 3528 GTCCGAGAAATGATCATACCGAATACACCTGCTCAAGGTGTGACCTGGGCAATGC 3587  
 QY 361 ACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 420  
 DB 3588 ACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 3647

QY 421 CATGGCTCCAGAGCTCCTGGAGGGCCAGGGGCTGTTCACAGACAGACATCTGGGCCAT 480  
 DB 3648 CATGGCTCCAGAGCTCCTGGAGGGCCAGGGGCTGTTCACAGACAGACATCTGGGCCAT 3707  
 QY 481 CGGTGTGACAGGCTTTCATCATCTGAGGCCCGAGTACCGGTGACAGCGAGGTGTCACG 540  
 DB 3708 CGGTGTGACAGGCTTTCATCATCTGAGGCCCGAGTACCGGTGACAGCGAGGTGTCACG 3767  
 QY 541 CGACCTGACAGAGACTGCGCAAGGGGCTGTGCGGCTGAGCCGCTGCTACGCGGGGCT 600  
 DB 3768 CGACCTGACAGAGACTGCGCAAGGGGCTGTGCGGCTGAGCCGCTGCTACGCGGGGCT 3827  
 QY 601 GTCCGGGGGCGCGGTGGCTTCTCGCAGACACTCTGTGCGCCCAAGCCCTGGGGCCGGCC 660  
 DB 3828 GTCCGGGGGCGCGGTGGCTTCTCGCAGACACTCTGTGCGCCCAAGCCCTGGGGCCGGCC 3887  
 QY 661 CTGCGGCTCCAGCTGCTGAGTCCCGTGGCTAAACAGAGAGGCGCCGCGCTGTTCGCG 720  
 DB 3888 CTGCGGCTCCAGCTGCTGAGTCCCGTGGCTAAACAGAGAGGCGCCGCGCTGTTCGCG 3947  
 QY 721 GCCCGCGCGCTGACCTTCCCTACCGCGCGGTGGCTCTTCGTGCGCAATCGCGAA 780  
 DB 3948 GCCCGCGCGCTGACCTTCCCTACCGCGCGGTGGCTCTTCGTGCGCAATCGCGAA 4007  
 QY 781 GAGACCGCGCTGCTGTATACAGAGGCAACAACCTGGCCAGGTGGCTGAGGTGCGCCCG 840  
 DB 4008 GAGACCGCGCTGCTGTATACAGAGGCAACAACCTGGCCAGGTGGCTGAGGTGCGCCCG 4067  
 QY 841 GCCACACCTTGGTCTCCCGCTGGGCTGCGCTGAGAGCGCCCAATAAAGCCCGCAGC 900  
 DB 4068 GCCACACCTTGGTCTCCCGCTGGGCTGCGCTGAGAGCGCCCAATAAAGCCCGCAGC 4127  
 QY 901 CGGGCGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 948  
 DB 4128 CGGGCGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 4175

RESULT 4  
 AAL43909  
 ID AAL43909 standard; DNA; 5207 BP.  
 XX  
 AC AAL43909;  
 XX  
 DT 19-SEP-2002 (first entry)  
 XX  
 DE Human kinase protein coding sequence 2.  
 XX  
 KW Human; gene; ds; Gene therapy; chromosome 1; kinase protein;  
 KW myosin light chain kinase subfamily; kinase protein-mediated disease;  
 KW transgenic animal.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 109..5106  
 FT /\*tag= a  
 FT /product= "Human kinase protein"  
 FT /replace(311, C)  
 FT /\*tag= c  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT /replace(311, T)  
 FT /\*tag= b  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT /replace(1741, T)  
 FT /\*tag= d  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT /replace(2714, C)  
 FT /\*tag= e  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT /replace(2745, T)  
 FT /\*tag= f  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT /replace(2859, A)  
 FT /variation



```
FT      /tag= g
FT      /standard_name= "single nucleotide polymorphism"
FT      variation :
FT      replace(3420, C)
FT      /tag= h
FT      /standard_name= "single nucleotide polymorphism"
XX      WO200240683-A2.
XX      23-MAY-2002.
XX      22-OCT-2001; 2001WO-US032616.
XX      14-NOV-2000; 2000US-00711134.
XX      17-MAY-2001; 2001US-00858664.
XX      (PEKE ) PE CORP NY.
XX      Wei M, Ketchum K, Di Francesco V, Beasley EM;
XX      WPI; 2002-500223/53.
XX      P-PSDB; AAO15372.
XX      New kinase proteins related to myosin light chain kinase subfamily and
XX      encoding polynucleotide, useful for diagnosing, treating disease or
XX      condition mediated by the kinase protein and for identifying modulators.
XX      Claim 23; Fig 3; 96pp; English.
XX      The invention comprises the amino acid and coding sequences (located on
XX      chromosome 1) of a human kinase protein that is related to the myosin
XX      light chain kinase subfamily. The human kinase DNA and protein sequences
XX      of the invention are useful for identifying agents that modulate the
XX      activity of the human kinase protein. Kinase-modulating agents are useful
XX      for treating a disease or condition mediated by a human kinase protein.
XX      The human kinase DNA sequences can be used to produce transgenic animals
XX      which are useful for studying the function of kinase proteins and
XX      identifying/evaluating modulators of kinase protein activity. The present
XX      DNA sequence encodes the human kinase protein of the invention.
XX      Sequence 5207 BP; 988 A; 1758 C; 1612 G; 849 T; 0 U; 0 Other;
XX      Query Match      95.4%; Score 926.8; DB 6; Length 5207;
XX      Best Local Similarity 99.8%; Pred. No. 1.9e-163;
XX      Matches 928; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX      QY      1 CAGCGGGGGGGCGCTGGCGCCAGATCATCCCTACACCCCAAGGACAGACAGCAGT 60
XX      DB      4278 CAGCGGGGGGGCGCTGGCGCCAGATCATCCCTACACCCCAAGGACAGACAGCAGT 4337
XX      QY      61 GCTGCGGAATACGAGGGCCCTCAAGGGCTGCGCCACCGCACCTGGCCAGCTGCACGC 120
XX      DB      4338 GCTGCGGAATACGAGGGCCCTCAAGGGCTGCGCCACCGCACCTGGCCAGCTGCACGC 4397
XX      QY      121 AGCTTACCTCAGCCCCCGGACCTGCTGCTATCTTGGAGCTGTGCTTGGCCCGGAGCT 180
XX      DB      4398 AGCTTACCTCAGCCCCCGGACCTGCTGCTATCTTGGAGCTGTGCTTGGCCCGGAGCT 4457
XX      QY      181 GCTCCCTGCTGGCGGAGAGGGCTCTCTACTCAGAACTCCGAGGTGAAGGACTACCTCTG 240
XX      DB      4458 GCTCCCTGCTGGCGGAGAGGGCTCTCTACTCAGAACTCCGAGGTGAAGGACTACCTCTG 4517
XX      QY      241 GCAGATGTTAGTGCCACCCAGTACCTGCACAAACAGACACATCTCTGACCTGGACCTGAG 300
XX      DB      4518 GCAGATGTTAGTGCCACCCAGTACCTGCACAAACAGACACATCTCTGACCTGGACCTGAG 4577
XX      QY      301 GTCGAGAAATGATCATCCGAATACAACTGCTCAAGTCTGAGCTGGACCTGGCAATGC 360
XX      DB      4578 GTCGAGAAATGATCATCCGAATACAACTGCTCAAGTCTGAGCTGGACCTGGCAATGC 4637
XX      QY      361 ACAGAGCTTCAGCCAGGAGAGGCTGCTGCCCTTCAGACAAAGTTCAAGGACTACCTAGAGAC 420
XX      DB      4638 ACAGAGCTTCAGCCAGGAGAGGCTGCTGCCCTTCAGACAAAGTTCAAGGACTACCTAGAGAC 4697
```

```
QY      421 CATGGCTCCAGAGCTCTCTGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 480
DB      4698 CATGGCTCCAGAGCTCTCTGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 4757
QY      481 CGGTGTGACAGCCTTCATCATGTCAGCCGCCAGTACCCGGTGAGCAGAGAGGTGCACG 540
DB      4758 CGGTGTGACAGCCTTCATCATGTCAGCCGCCAGTACCCGGTGAGCAGAGAGGTGCACG 4817
QY      541 CGACCTGTCAGAGAGGACTCGCCAAAGGGCTGGTCCGGCTGAGCCCGCTGTACCGGGCT 600
DB      4818 CGACCTGTCAGAGAGGACTCGCCAAAGGGCTGGTCCGGCTGAGCCCGCTGTACCGGGCT 4877
QY      601 GTCCGGGGGGCGGCTGCTTCTTCCGCGACGACTCTGTGCGCCAGCCCTGGGGCGGGCC 660
DB      4878 GTCCGGGGGGCGGCTGCTTCTTCCGCGACGACTCTGTGCGCCAGCCCTGGGGCGGGCC 4937
QY      661 CTGCGGCTCCAGCTGCTGCTGTCAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
DB      4938 CTGCGGCTCCAGCTGCTGCTGTCAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4997
QY      721 GCCCGCGCGCTGACCTTCCCTACCGCGGGCTGCGCGTCTTCGTGCGCAATCGCGAGAA 780
DB      4998 GCCCGCGCGCTGACCTTCCCTACCGCGGGCTGCGCGTCTTCGTGCGCAATCGCGAGAA 5057
QY      781 GAGACCGCGCTGCTGTACAAGAGGCAACAACCTGGCCAGGTGCGCTGAGGGTTCGCCCG 840
DB      5058 GAGACCGCGCTGCTGTACAAGAGGCAACAACCTGGCCAGGTGCGCTGAGGGTTCGCCCG 5117
QY      841 GCCACACCTTGTCTCCCGCTGGGGTCTGCTGTCAGACGCGCCAAATAAAACGCCACAGC 900
DB      5118 GCCACACCTTGTCTCCCGCTGGGGTCTGCTGTCAGACGCGCCAAATAAAACGCCACAGC 5177
QY      901 CGGCGGAGAAAAAATAAAAAAATAAAAAA 930
DB      5178 CGGCGGAGAAAAAATAAAAAAATAAAAAA 5207
XX      RESULT 5
XX      AAL43908
XX      ID AAL43908 standard; DNA; 5207 BP.
XX      AC AAL43908;
XX      DT 19-SEP-2002 (first entry)
XX      DE Human kinase protein coding sequence 1.
XX      KW Human; gene; ds; gene therapy; chromosome 1; kinase protein;
XX      KW myosin light chain kinase subfamily; kinase protein-mediated disease;
XX      KW transgenic animal.
XX      OS Homo sapiens.
XX      FH Key Location/Qualifiers
XX      CDS 109..5106
XX      FT /*tag= a
XX      FT /product= "Human kinase protein"
XX      PN WO200240683-A2.
XX      PD 23-MAY-2002.
XX      XX 22-OCT-2001; 2001WO-US032616.
XX      PF 14-NOV-2000; 2000US-00711134.
XX      PR 17-MAY-2001; 2001US-00858664.
XX      XX (PEKE ) PE CORP NY.
XX      XX Wei M, Ketchum K, Di Francesco V, Beasley EM;
XX      PI WPI; 2002-500223/53.
XX      DR P-PSDB; AAO15372.
XX      DR
```

XX New kinase proteins related to myosin light chain kinase subfamily and  
 PT encoding polynucleotide, useful for diagnosing, treating disease or  
 PT condition mediated by the kinase protein and for identifying modulators.

XX Claim 23; Fig 1; 96pp; English.

XX The invention comprises the amino acid and coding sequences (located on  
 CC chromosome 1) of a human kinase protein that is related to the myosin  
 CC light chain kinase subfamily. The human kinase DNA and protein sequences  
 CC of the invention are useful for identifying agents that modulate the  
 CC activity of the human kinase protein. Kinase-modulating agents are useful  
 CC for treating a disease or condition mediated by a human kinase protein.  
 CC The human kinase DNA sequences can be used to produce transgenic animals  
 CC which are useful for studying the function of kinase proteins and  
 CC identifying/evaluating modulators of kinase protein activity. The present  
 CC DNA sequence encodes the human kinase protein of the invention

XX Sequence 5207 BP; 988 A; 1758 C; 1612 G; 849 T; 0 U; 0 Other;

Query Match 95.4%; Score 926.8; DB 6; Length 5207;  
 Best Local Similarity 99.8%; Pred. No. 1.9e-163;  
 Matches 928; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGCGGGCGGCGCTGGCGGCGGAGATCATCCCTACCCACCCCAAGCAGACAGCAGT 60  
 DB 4278 CAGCGGGCGGCGCTGGCGGCGGAGATCATCCCTACCCACCCCAAGCAGACAGCAGT 4337  
 QY 61 GGTGGCGGAATACAGAGGCGCTCAAGGGGCTGCGGCACCCGACCTGGCCAGCTGCACGC 120  
 DB 4338 GGTGGCGGAATACAGAGGCGCTCAAGGGGCTGCGGCACCCGACCTGGCCAGCTGCACGC 4397  
 QY 121 AGCTTACTAGCCCCCGGACCTGGTGCTCATCTTGGAGCTGTGCTCTGGGCGGAGCT 180  
 DB 4398 AGCTTACTAGCCCCCGGACCTGGTGCTCATCTTGGAGCTGTGCTCTGGGCGGAGCT 4457  
 QY 181 GTCCCTCCCTGCTGGCGGAGAGGGCTCTCTACTCAGAAATCCGAGGTGAAGGACTACCTGTG 240  
 DB 4458 GTCCCTCCCTGCTGGCGGAGAGGGCTCTCTACTCAGAAATCTGAGTGAAGGACTACCTGTG 4517  
 QY 241 CGAGATGTAGTGCACCCAGTACCTGCAACACAGACATCTCTGACCTGGACCTGAG 300  
 DB 4518 GCAGATGTAGTGCACCCAGTACCTGCAACACAGACATCTCTGACCTGGACCTGAG 4577  
 QY 301 GTCCGAGAACATGATCATCACCAGATACACCTGCTCAAGGTCTGAGCTGGGCAATGC 360  
 DB 4578 GTCCGAGAACATGATCATCACCAGATACACCTGCTCAAGGTCTGAGCTGGGCAATGC 4637  
 QY 361 ACAGAGCTTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 420  
 DB 4638 ACAGAGCTTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 4697  
 QY 421 CATGGCTCCAGACTCTCGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCCAT 480  
 DB 4698 CATGGCTCCAGACTCTCGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCCAT 4757  
 QY 481 CGGTGTGACAGCTTCATCATGCTGAGCGCCGAGTACCCGCTGAGCAGCGAGGCTGCACG 540  
 DB 4758 CGGTGTGACAGCTTCATCATGCTGAGCGCCGAGTACCCGCTGAGCAGCGAGGCTGCACG 4817  
 QY 541 CGACTGTCAGAGAGACTGCGCAAGGGGCTGGTCCGGCTGAGCGCTGCTACCGGGGCT 600  
 DB 4818 CGACTGTCAGAGAGACTGCGCAAGGGGCTGGTCCGGCTGAGCGCTGCTACCGGGGCT 4877  
 QY 601 GTCCGGGGGCGCGCTGCTTCCTCGCAGACATCTGTGGCGCCAGCCCTGGGCGCGGCC 660  
 DB 4878 GTCCGGGGGCGCGCTGCTTCCTCGCAGACATCTGTGGCGCCAGCCCTGGGCGCGGCC 4937  
 QY 661 GTGGCGCTCAGCTGCTGAGTGCCTGGTGGCTTAACAGAGAGGGGCCCGCTGTCGG 720  
 DB 4938 GTGGCGCTCAGCTGCTGAGTGCCTGGTGGCTTAACAGAGAGGGGCCCGCTGTCGG 4997  
 QY 721 GCCCGCGCGCGTACCTTCCTACCGCGGCTGCGGCTCTTCGTGCGCAATCGCGAGAA 780

DB 4998 GCCCGCGCGCGTGGACCTTCCTACCGCGGCTGGCGCTTCGTGCGCAATCGCGAGAA 5057  
 QY 781 GAGACGCGCGCTGCTGTACAGAGGACACACCTGGCCAGGTCGCTGAGGTCGCCCG 840  
 DB 5058 GAGACGCGCGCTGCTGTACAGAGGACACACCTGGCCAGGTCGCTGAGGTCGCCCG 5117  
 QY 841 GCCACACCTTGGTCTCTCCCGCTGGGGTGGCTGAGACGCGCCCAATAAAGCGCCAGC 900  
 DB 5118 GCCACACCTTGGTCTCTCCCGCTGGGGTGGCTGAGACGCGCCCAATAAAGCGCAGC 5177  
 QY 901 CGGGCGAGAAAAAATAAATAAATAAATAAATAA 930  
 DB 5178 CGGGCGAGAAAAAATAAATAAATAAATAAATAA 5207

## RESULT 6

AAD26467

ID AAD26467 standard; cDNA; 3225 BP.

XX

AC AAD26467;

XX

DT 26-MAR-2002 (first entry)

XX

DE Human kinase PKIN-20 cDNA.

XX

KW Human; kinase; PKIN-20; cancer; leukaemia; adenocarcinoma; osteoporosis;  
 KW immune disorder; atherosclerosis; Crohn's disease; Hodgkin's disease;  
 KW Acquired Immune Deficiency Syndrome; AIDS; Addison's disease; anaemia;  
 KW allergy; asthma; adult respiratory distress syndrome; multiple sclerosis;  
 KW autoimmune thyroiditis; bronchitis; diabetes mellitus; osteoarthritis;  
 KW Good pasture's syndrome; Graves' disease; pancreatitis; psoriasis;  
 KW rheumatoid arthritis; ulcerative colitis; cirrhosis; Cushing's syndrome;  
 KW hepatitis; hypothyroidism; cerebral palsy; cataract; angina pectoris;  
 KW cardiovascular disease; hypertension; vasculitis; myocarditis; obesity;  
 KW congestive heart failure; ischaemic heart disease; lung tumour; gout;  
 KW fatty liver; Niemann-Pick's disease; gene therapy; ss.

XX Homo sapiens.

OS

Key Location/Qualifiers

CDS 526..3141

FT /\*tag= a

FT /product= "Human PKIN-20 protein"

XX

PN WC200196547-A2.

XX

PD 20-DEC-2001.

XX

PF 14-JUN-2001; 2001WO-US019444.

XX

PP 15-JUN-2000; 2000US-0212073P.

XX

PR 23-JUN-2000; 2000US-0213467P.

XX

PR 30-JUN-2000; 2000US-0215651P.

XX

PR 07-JUL-2000; 2000US-0216605P.

XX

PR 13-JUL-2000; 2000US-0218372P.

XX

PR 25-AUG-2000; 2000US-0228056P.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

Yue H, Lal P, Bandman O, Borowsky ML, Au-Young J, Lu Y;

PI Gandhi AR, Tribouley CM, Walla NK, Yao MG, Lu DAM, Greenwald SR;

PI Ramkumar J, Griffin JA, Kearney L, Burford N, Nguyen DB, Tang YT;

PI Baughn MR, He A, Thornton M, Rafalia A, Patterson C, Gururajan R;

PI Lo TP, Khan F, Recipon SA, Azimzai Y, Policky JL, Ding L;

PI Grether M, Elliott VS, Thangavelu K, Batra S, Ison CH;

XX WPI; 2002-090207/12.

DR P-PSDB; AAE16274.

XX

New polypeptides, useful for diagnosing, treating or preventing disorders  
 of growth and development, cardiovascular and lipid, and diseases such as  
 cancer, comprise human kinase polypeptides.

Claim 5; Page 191; 197pp; English.

The invention relates to human kinase PKIN proteins and their corresponding cDNAs. A composition containing PKIN agonist is useful for treating a disease or condition associated with decreased expression of PKIN and a composition comprising PKIN antagonist is useful for treating a disease or condition associated with overexpression of PKIN. The disorders include cancer (leukaemia, adenocarcinoma, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, Hodgkin's disease); immune disorder (Acquired Immune Deficiency Syndrome (AIDS), asthma, Addison's disease, atherosclerosis, anaemia, allergies, adult respiratory distress syndrome, autoimmune thyroiditis, gout, bronchitis, Crohn's disease, diabetes mellitus, multiple sclerosis, Good pasture's syndrome, Graves' disease, osteoarthritis, osteoporosis, pancreatitis, psoriasis, Reiter's syndrome, rheumatoid arthritis, Sjogren's syndrome, uveitis, ulcerative colitis, bacterial, parasitic, fungal, viral, protozoal and helminthic infections), growth and development disorders (arteriosclerosis, cirrhosis, hepatitis, Cushing's syndrome, hypothyroidism, cerebral palsy, cataracts), cardiac vascular disease (arteriovenous fistula, hypertension, vasculitis, aneurysms, congestive heart failure, angina pectoris, myocarditis, ischaemic heart disease, chronic bronchitis, lung tumours), lipid disorder (fatty liver, Fabry's disease, Niemann-Pick's disease, hypocholesterolaemia, obesity). PKIN DNA is useful for assessing toxicity of a test compound and in gene therapy. The present sequence is human PKIN-20 cDNA

Sequence	3225 BP; 596 A; 1104 C; 1017 G; 508 T; 0 U; 0 Other;
Query Match	93.9%; Score 911.4; DB 6; Length 3225;
Best Local Similarity	99.9%; Pred. No. 1.3e-160;
Matches 912; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 CAGGGGGGGCGCTGGCGCCCAAGATCATCCCTACACCCCAAGACACAGACAGT 60
DB	
2313	CAGGGGGGGCGCTGGCGCCCAAGATCATCCCTACACCCCAAGACACAGACAGT 2372
QY	61 GCTGCGCGAATACAGAGCCCTCAAGGGCTGCGGCCACCGCACCTGCGCCAGCTGCACGC 120
DB	
2373	GCTGCGCGAATACAGAGCCCTCAAGGGCTGCGGCCACCGCACCTGCGCCAGCTGCACGC 2432
QY	121 AGCCTACTCAGCCCCCGGCACCTGTGTGCTCATCTTGGAGCTGTGCTCTGGGCCCGAGCT 180
DB	
2433	AGCCTACTCAGCCCCCGGCACCTGTGTGCTCATCTTGGAGCTGTGCTCTGGGCCCGAGCT 2492
QY	181 GCTCCCTGCTGCGCCGAGAGGGGCTCTCTACTCAGATCCGAGGTGAAGACTACCTGTG 240
DB	
2493	GCTCCCTGCTGCGCCGAGAGGGGCTCTCTACTCAGATCCGAGGTGAAGACTACCTGTG 2552
QY	241 GCAGATGTTGAGTGCCACCCAGTACCTGCACACCCAGCACATCTGCACCTGGACCTGAG 300
DB	
2553	GCAGATGTTGAGTGCCACCCAGTACCTGCACACCCAGCACATCTGCACCTGGACCTGAG 2612
QY	301 GTCCGAAACATGATCATCACCGAATACACCTGCTCAAGGTCTGAGACCTGGGCAATGC 360
DB	
2613	GTCCGAAACATGATCATCACCGAATACACCTGCTCAAGGTCTGAGACCTGGGCAATGC 2672
QY	361 ACAGAGCCTCAGCCAGAGAGAGGTGCTGCCCTCAGACAAGTTCAGAGACTACCTAGAGAC 420
DB	
2673	ACAGAGCCTCAGCCAGAGAGAGGTGCTGCCCTCAGACAAGTTCAGAGACTACCTAGAGAC 2732
QY	421 CATGGCTCCAGAGCTCTCTGAGGGGCCAGGGGGCTGTTCCACAGACACATCTGGGCCAT 480
DB	
2733	CATGGCTCCAGAGCTCTCTGAGGGGCCAGGGGGCTGTTCCACAGACACATCTGGGCCAT 2792
QY	481 CGGTGTGCACAGCCTTCATCTGCTGAGCGCGCGAGTACCCGGTGTAGCAGCGAGGCTGCACG 540
DB	
2793	CGGTGTGCACAGCCTTCATCTGCTGAGCGCGCGAGTACCCGGTGTAGCAGCGAGGCTGCACG 2852
QY	541 CGACCTGCAGAGAGGACTGCGCAAGGGGCTGGTCCGGCTGAGCCGCTGTACTACGGGGGCT 600
DB	
2853	CGACCTGCAGAGAGGACTGCGCAAGGGGCTGGTCCGGCTGAGCCGCTGTACTACGGGGGCT 2912

Qy	601	GTCCGGGGCGCGTGGCCTTCTCTGGCAGCACTCTGTGCGCCACGACCCCTGGGGCCGGCC	660
Db	2913	GTCCGGGGCGCGTGGCCTTCTCTGGCAGCACTCTGTGCGCCACGACCCCTGGGGCCGGCC	2972
Qy	661	CTGCGGCTCCAGCTGCCTGCAGTGCCTGGCTAAACAGAGAGAGGGCCCGCCCTGTTTCGCG	720
Db	2973	CTGCGGCTCCAGCTGCCTGCAGTGCCTGGCTAAACAGAGAGAGGGCCCGCCCTGTTTCGCG	3032
Qy	721	GCCCGGCCCGTGCACCTTCCCTACCGCGCGGCTGCGGCTTCTGTCGCGCAATCGCGAGAA	780
Db	3033	GCCCGGCCCGTGCACCTTCCCTACCGCGCGGCTGCGGCTTCTGTCGCGCAATCGCGAGAA	3092
Qy	781	GAGACGGCGCTGCTGTACAGAGGACACCTGGCCCGGCTGCGCTGAGGGTGCGCCCG	840
Db	3093	GAGACGGCGCTGCTGTACAGAGGACACCTGGCCCGGCTGCGCTGAGGGTGCGCCCG	3152
Qy	841	GCCACACCTTGGTCTCTCCCGCTGGGGGTCGCTGCAGACGCGCAATAAAAAACGCCAGC	900
Db	3153	GCCACACCTTGGTCTCTCCCGCTGGGGGTCGCTGCAGACGCGCAATAAAAAACGCCAGC	3212
Qy	901	CGGGCGAGAAAA 913	
Db	3213	CGGGCGAGAAAA 3225	
RESULT 7			
AAD38865			
ID	AAD38865 standard; cDNA; 5454 BP.		
XX			
AC	AAD38865;		
XX			
DT	23-SEP-2002 (first entry)		
XX			
DE	Human kinase (PKIN)-22 cDNA.		
XX			
KW	Human; kinase; PKIN; cancer; immune system disorder; atherosclerosis;		
KW	acquired immune deficiency syndrome; AIDS; Addison's disease; allergy;		
KW	asthma; multiple sclerosis; psoriasis; arteriosclerosis; cirrhosis;		
KW	development; hepatitis; cardiovascular; hypertension; drug screening;		
KW	myocardial infarction; Goodpasture's syndrome; lipid disorder; growth;		
KW	fatty liver; Gaucher's disease; Niemann-Pick's disease; anorectic;		
KW	hypercholesterolaemia; obesity; Gene therapy; cytostatic; anti-HIV;		
KW	neuroprotective; hepatotropic; hypotensive; cardiac; nephrotropic;		
XX	hyperlipidaemia; enzyme; gene; ss.		
XX			
OS	Homo sapiens.		
XX			
PH	Key Location/Qualifiers		
FT	CDS	373..5370	
FT	/*tag= a		
FT	/product= "Human kinase (PKIN)-22"		
XX			
PN	WO200233099-A2.		
XX			
PD	25-APR-2002.		
XX			
PF	20-OCT-2001; 2001WO-US047728.		
XX			
PR	20-OCT-2000; 2000US-0242410P.		
PR	27-OCT-2000; 2000US-0244068P.		
PR	03-NOV-2000; 2000US-0245708P.		
PR	09-NOV-2000; 2000US-024672P.		
PR	16-NOV-2000; 2000US-0249565P.		
PR	22-NOV-2000; 2000US-0252730P.		
PR	01-DEC-2000; 2000US-0250807P.		
XX			
PA	(INCY-) INCYTE GENOMICS INC.		
XX			
PI	Gururajan R, Baughn WR, Walia NK, Elliott VS, Xu Y, Arvizu C;		
PI	Yao MG, Ramkumar J, Ding L, Tang YT, Hafalia AJA, Nguyen DB;		
PI	Gandhi AR, Lu Y, Yue H, Burford N, Bandman O, Tribouley CM, Lal PG;		
PI	Recipon SA, Lu DAM, Borowsky ML, Thornton M, Swarnaker A;		
PI	Thangavelu K, Khan FA, Ison CH;		

XX WPI; 2002-454603/48.  
 DR P-PSDB; AAE24151.  
 XX  
 PT New human kinase polypeptide, for diagnosing, preventing and treating  
 PT cancer, immune system disorders, growth and development disorders,  
 PT cardiovascular disorders and lipid disorders.  
 XX  
 PS Claim 5; Page 209-210; 210pp; English.  
 XX  
 CC The invention relates human kinases (PKIN) and their corresponding  
 CC nucleic acid sequences. PKIN and its DNA are useful for diagnosing,  
 CC treating and preventing cancer, an immune system disorder (e.g., acquired  
 CC immune deficiency syndrome (AIDS), Addison's disease, allergy, asthma,  
 CC atherosclerosis, multiple sclerosis, psoriasis), disorders affecting  
 CC growth and development (e.g., arteriosclerosis, cirrhosis, hepatitis),  
 CC cardiovascular disorder (e.g., hypertension, myocardial infarction,  
 CC Goodpasture's syndrome), and a lipid disorder (e.g., fatty liver,  
 CC Gaucher's disease, Niemann-Pick's disease, hypercholesterolaemia,  
 CC hyperlipidaemia, obesity), and for assessing the effects of exogenous  
 CC compounds. Anti-PKIN antibody is useful in a diagnostic test for a  
 CC condition or a disease associated with the expression of PKIN in a  
 CC biological sample. A composition comprising PKIN or an agonist or  
 CC antagonist of PKIN is useful for treating a disease or condition  
 CC associated with decreased or increased expression of functional PKIN.  
 CC PKIN is useful in a number of drug screening techniques and to analyse  
 CC the proteome of a tissue or cell type. PKIN DNA is useful for creating  
 CC knockin humanised animals or transgenic animals to model human diseases,  
 CC and in somatic or germline gene therapy. The present sequence is human  
 CC PKIN cDNA  
 XX  
 SQ Sequence 5454 BP; 1019 A; 1853 C; 1686 G; 896 T; 0 U; 0 Other;

Query Match 93.9%; Score 911.4; DB 6; Length 5454;  
 Best Local Similarity 99.9%; Pred. No. 1.4e-160;  
 Matches 912; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGGGCGGCGCTGGCGGCGAGATCATCCCTACCCACCCCAAGGACAGACAGT 60  
 DB 4542 CAGCGGGCGGCGCTGGCGGCGAGATCATCCCTACCCACCCCAAGGACAGACAGT 4601  
 QY 61 GCTGGCGGATACAGAGGCGCTCAAGGGCGCTGGCGGCGAGTCCGCGGCGAGTGCAGCG 120  
 DB 4602 GCTGGCGGATACAGAGGCGCTCAAGGGCGCTGGCGGCGAGTCCGCGGCGAGTGCAGCG 4661  
 QY 121 AGCTTACTCAGCCCCCGGCGACCTGTGTCTCTTGGAGCTGTGTCTGGGCGCGAGCT 180  
 DB 4662 AGCTTACTCAGCCCCCGGCGACCTGTGTCTCTTGGAGCTGTGTCTGGGCGCGAGCT 4721  
 QY 181 GCTCCCTGCTGCGCGAGGCGCTCTACTCAGATCCGAGGTGAAGGACTACCTGTG 240  
 DB 4722 GCTCCCTGCTGCGCGAGGCGCTCTACTCAGATCCGAGGTGAAGGACTACCTGTG 4781  
 QY 241 GCAGATGTGATGCGACCCAGTACCTGCGACACACACATCTCCACCTGACCTGAG 300  
 DB 4782 GCAGATGTGATGCGACCCAGTACCTGCGACACACACATCTCCACCTGACCTGAG 4841  
 QY 301 GTCCGAGAACATGATCATCACCGGAATACACCTGTCTCAAGTCTGTGACCTGGGCAATGC 360  
 DB 4842 GTCCGAGAACATGATCATCACCGGAATACACCTGTCTCAAGTCTGTGACCTGGGCAATGC 4901  
 QY 361 ACAGAGCTCAGCCAGAGAGGTGCTGCCCTCAGACAACTTCAAGGACTACCTAGAGAC 420  
 DB 4902 ACAGAGCTCAGCCAGAGAGGTGCTGCCCTCAGACAACTTCAAGGACTACCTAGAGAC 4961  
 QY 421 CATGGCTCCAGAGCTCTCTGGAGGCGCCAGGGGCTGTTCACACAGACAGACATCTGGGCCAT 480  
 DB 4962 CATGGCTCCAGAGCTCTCTGGAGGCGCCAGGGGCTGTTCACACAGACAGACATCTGGGCCAT 5021  
 QY 481 CGGTGTGACAGCTTCTCATCTGCTGAGCGCGGAGTACCCGGTGAGCAGCGAGGCTGCAGC 540  
 DB 5022 CGGTGTGACAGCTTCTCATCTGCTGAGCGCGGAGTACCCGGTGAGCAGCGAGGCTGCAGC 5081

QY 541 CGACCTGCAGAGAGACTGCGCAAGGGGCTGGTCCGGCTGAGCGGCTGCTACGCGGGGCT 600  
 DB 5082 CGACCTGCAGAGAGACTGCGCAAGGGGCTGGTCCGGCTGAGCGGCTGCTACGCGGGGCT 5141  
 QY 601 GTCGGGGGGCGCGGCTTCCTGCGCAGCACTCTGTGCGCCCGCAGCCCTGGGGCCGCGCC 660  
 DB 5142 GTCGGGGGGCGCGGCTTCCTGCGCAGCACTCTGTGCGCCCGCAGCCCTGGGGCCGCGCC 5201  
 QY 661 CTGCGCGTCCAGCTGCTGCACTGCGCGTAAACAGAGAGGCGCGCGGCTGTCTGCGG 720  
 DB 5202 CTGCGCGTCCAGCTGCTGCACTGCGCGTAAACAGAGAGGCGCGCGGCTGTCTGCGG 5261  
 QY 721 GCCCGCGCGGCTGACCTTCCTACCGCGCGCTGGCGGCTGCTGCGCAATCGCGAGAA 780  
 DB 5262 GCCCGCGCGGCTGACCTTCCTACCGCGCGCTGGCGGCTGCTGCGCAATCGCGAGAA 5321  
 QY 781 GAGACGCGGCTGCTGTATACAGAGGCGCAACCTGGCCCGAGTGGCTGAGGGTGGCCCGG 840  
 DB 5322 GAGACGCGGCTGCTGTATACAGAGGCGCAACCTGGCCCGAGTGGCTGAGGGTGGCCCGG 5381  
 QY 841 GCCACACCTTGGTCTCCCGCTGGGGGTGCTGCGACGCGCGCAATAAAGCCCGCAGC 900  
 DB 5382 GCCACACCTTGGTCTCCCGCTGGGGGTGCTGCGACGCGCGCAATAAAGCCCGCAGC 5441  
 QY 901 CGGGCGAGAAAAA 913  
 DB 5442 CGGGCGAGAAAAA 5454

RESULT 8  
 AAH46904  
 ID AAH46904 standard; cDNA; 4936 BP.  
 XX AC AAH46904;  
 XX DT 25-SEP-2001 (first entry)  
 XX DE cDNA encoding human protein kinase SGK145.  
 XX KW Protein kinase; enzyme; cytostatic; nootropic; neuroprotective; human;  
 KW antiparkinsonian; virucide; antibacterial; antifungal; antimigraine;  
 KW analgesic; hypotensive; hypertensive; immunosuppressive; antiallergic;  
 KW antipsoriatic; antirheumatic; antiarthritic; ophthalmological; anorectic;  
 KW osteopathic; thrombolytic; antiarteriosclerotic; antiasthmatic;  
 KW vasotropic; antidiabetic; gene therapy; ss.  
 XX OS Homo sapiens.  
 XX FH Key  
 FT CDS Location/Qualifiers  
 FT 1..4851  
 FT /\*tag= a  
 XX WO200155356-A2.  
 XX PD 02-AUG-2001.  
 XX PF 25-JAN-2001; 2001WO-US002337.  
 XX PR 25-JAN-2000; 2000US-0178078P.  
 PR 31-JAN-2000; 2000US-0179364P.  
 PR 17-FEB-2000; 2000US-0183173P.  
 PR 17-MAR-2000; 2000US-0190162P.  
 PR 29-MAR-2000; 2000US-0193404P.  
 PR 13-NOV-2000; 2000US-0247013P.  
 XX (SUGEN-) SUGEN INC.  
 XX PA Plowman G, Whyte D, Manning G, Sudarsanam S, Martinez R;  
 XX WPI; 2001-476202/51.  
 DR P-PSDB; AAB85504.  
 XX Kinase polypeptides useful for treating cancers, Alzheimer's disease,

PT viral infections, diabetes, obesity, organ transplant rejection and  
PT rheumatoid arthritis.  
PS Example 1; Page 204-205; 218pp; English.  
XX  
CC The invention provides human protein kinases and protein kinase-like  
CC enzymes and polynucleotides encoding the polypeptides. The kinase  
CC polypeptides and their modulators are useful for treating a disease or  
CC disorder such as cancer, immune-related diseases, cardiovascular disease,  
CC brain or neuronal-associated disease and metabolic disorders, including  
CC cancers of tissues, cancers of hematopoietic origin, diseases of the  
CC central nervous system, diseases of the peripheral nervous system,  
CC Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic  
CC lateral sclerosis, viral infections, infections caused by prions,  
CC bacteria and fungi, ocular diseases, migraines, pain, sexual dysfunction,  
CC mood disorders, attention disorders, cognition disorders, hypotension,  
CC hyperextension, psychotic disorders, neurological disorders, dyskinesias,  
CC metabolic disorders, and organ transplant rejection. They are also useful  
CC for treating rhinitis, autoimmunity, atherosclerosis, psoriasis,  
CC osteoarthritis, asthma, chronic inflammatory pelvic disease, chronic  
CC inflammatory bowel disease, rheumatoid arthritis, metabolic disorders  
CC such as diabetes, obesity, cardiovascular diseases such as reperfusion  
CC injury, coronary thrombosis, clotting disorders and atherosclerosis,  
CC ocular diseases such as glaucoma, retinopathy and macular degeneration,  
CC psychiatric and neurological disorders such as anxiety, schizophrenia,  
CC dementia, manic depression, etc. The polynucleotides are useful in gene  
CC therapy techniques to treat the above mentioned disorders. Sequences  
CC AAH46831-46922 represent human protein kinases encoding cDNA molecules  
XX  
SQ Sequence 4936 BP; 923 A; 1695 C; 1524 G; 794 T; 0 U; 0 Other;

Query Match 93.3%; Score 906.4; DB 4; Length 4936;  
Best Local Similarity 99.9%; Pred. NO. 1.2e-159;  
Matches 907; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CAGCGGGGGCGCTGGCGCCAGATCATCCCTACACCCCAAGACAGACAGCAGT 60  
DB 4029 CAGCGGGGGCGCTGGCGCCAGATCATCCCTACACCCCAAGACAGACAGCAGT 4088  
QY 61 GCTGCGGAATACGAGGCGCTTAAGGGCTTGGCCACCCGCACTGCGCCAGCTGCACG 120  
DB 4089 GCTGCGGAATACGAGGCGCTTAAGGGCTTGGCCACCCGCACTGCGCCAGCTGCACG 4148  
QY 121 AGCTTACCTCAGCCCGCCGACCTGGTCTCATCTTGGAGCTGTCTTGGCCCGAGCT 180  
DB 4149 AGCTTACCTCAGCCCGCCGACCTGGTCTCATCTTGGAGCTGTCTTGGCCCGAGCT 4208  
QY 181 GCTCCCTGCTGGCCGAGAGGCGCTCCTACTCAGAAATCCGAGGTGAAGACTACCTGTG 240  
DB 4209 GCTCCCTGCTGGCCGAGAGGCGCTCCTACTCAGAAATCCGAGGTGAAGACTACCTGTG 4268  
QY 241 GCAGATGTTGAGTGCACCCAGTACTCGCAACACGACATCTCGACCTGGACCTGAG 300  
DB 4269 GCAGATGTTGAGTGCACCCAGTACTCGCAACACGACATCTCGACCTGGACCTGAG 4328  
QY 301 GTCCGAGAACATGATCATCACCGAATACAACTGCTCAAGGTCTGGACCTGGGCAATGC 360  
DB 4329 GTCCGAGAACATGATCATCACCGAATACAACTGCTCAAGGTCTGGACCTGGGCAATGC 4388  
QY 361 ACAGAGCTTCAGCAGGAGAGGTGCTCCCTCAGCAAGTTCAAGGACTACCTAGAGAC 420  
DB 4389 ACAGAGCTTCAGCAGGAGAGGTGCTCCCTCAGCAAGTTCAAGGACTACCTAGAGAC 4448  
QY 421 CATGGCTCCAGAGCTTCCTGGAGGGCCAGGGGCTCTTCCACAGACAGACATCTGGGCCAT 480  
DB 4449 CATGGCTCCAGAGCTTCCTGGAGGGCCAGGGGCTCTTCCACAGACAGACATCTGGGCCAT 4508  
QY 481 CGGTGTGACAGCCTTTCATCATCTGAGCGCCGAGTACCCGGTGAACAGCGAGGTGCACG 540  
DB 4509 CGGTGTGACAGCCTTTCATCATCTGAGCGCCGAGTACCCGGTGAACAGCGAGGTGCACG 4568  
QY 541 CCACCTGCAGAGAGACTCGCAAGGGCTGTCTCGGCTGAGCCCTGCTACCGGGGCT 600

DB 4569 CGACCTGCAGAGAGGACTGCGCAAGGGGCTGTCTCCGGTGTAGCCCTGCTACGCGGGGT 4628  
QY 601 GTCCGGGGGGCGCGCTGGCTTCTCTGCGCAGACACTGTGCGCCAGCCCTTGGGGCGCGGC 660  
DB 4629 GTCCGGGGGGCGCGCTGGCTTCTCTGCGCAGACACTGTGCGCCAGCCCTTGGGGCGCGGC 4688  
QY 661 CTGCGCGTCCAGCTGCTGCTGCACTGCGCCGCTTAACAGAGAGGCGCCGCGCTGTTCGGC 720  
DB 4689 CTGCGCGTCCAGCTGCTGCTGCACTGCGCCGCTTAACAGAGAGGCGCCGCGCTGTTCGGC 4748  
QY 721 GCGCGCGCGCTGACTTCTTCCCTACCGCGGCTGCGCGCTTCTTCTGCGCAATCGCGAGAA 780  
DB 4749 GCGCGCGCGCTGACTTCTTCCCTACCGCGGCTGCGCGCTTCTTCTGCGCAATCGCGAGAA 4808  
QY 781 GAGACGCGCGCTGCTGTATCAAGAGGACAACTGCGCCAGGTGCGCTGAGGGTTCGCCCCG 840  
DB 4809 GAGACGCGCGCTGCTGTATCAAGAGGACAACTGCGCCAGGTGCGCTGAGGGTTCGCCCCG 4868  
QY 841 GCGACACCTTCTGCTTCCCTGCGCGGCTGCGCGCTTCTTCTGCGCAATCGCGAGC 900  
DB 4869 GCGACACCTTCTGCTTCCCTGCGCGGCTGCGCGCTTCTTCTGCGCAATCGCGAGC 4928  
QY 901 CGGGCGAG 908  
DB 4929 CGGGCGAG 4936  
RESULT 9  
AAC62286  
ID AAC62286 standard; cDNA; 5007 BP.  
XX  
AC AAC62286;  
XT  
DT 19-MAR-2001 (first entry)  
XX  
DE cDNA encoding a full length human signal transduction polypeptide.  
XX  
KW Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;  
KW congestive heart failure; dilated congestive cardiomyopathy;  
KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;  
KW mitral valve disease; aortic valve disease; tricuspid valve disease;  
KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;  
KW atherosclerosis; cardiac tumour; microbial infection; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
CDS 97..4929  
FT /tag= a  
FT /product= "signal transduction polypeptide H19G5"  
XX  
PN WO200063381-A1.  
XX  
PD 26-OCT-2000.  
XX  
PF 11-APR-2000; 2000MO-US009488.  
XX  
PR 16-APR-1999; 99US-0129553P.  
XX  
PA (SCIO-) SCIOS INC.  
XX  
PI Zeng W, Stanton L, Kong H;  
XX  
PI WPI; 2001-007013/01.  
XX  
DR P-PSDB; AAB30568.  
XX  
PT Novel h19G5 polypeptides capable of regulating signal transduction and  
PT exhibiting kinase activity useful for identifying antibodies to treat  
PT cardiac diseases, and additional mediators of signal transduction.  
XX  
PS Claim 4; Page 59-61; 81pp; English.  
XX  
CC The present sequence encodes a human protein with putative function in

CC signal transduction. The polypeptide is designated H19G5. The protein is  
 CC capable of regulating signal transduction and exhibits kinase activity.  
 CC The H19G5 transcript is expressed in the heart. H19G5 polypeptides and  
 CC polynucleotides are useful for preventing or treating a cardiac disease,  
 CC such as congestive heart failure, dilated congestive cardiomyopathy,  
 CC hypertrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve  
 CC disease, aortic valve disease or tricuspid valve disease, angina  
 CC pectoris, myocardial infarction, cardiac arrhythmia, pulmonary, arterial  
 CC or renovascular hypertension, arteriosclerosis, atherosclerosis and  
 CC cardiac tumours in humans. The polypeptide is also useful for detecting  
 CC the expression of a protein capable of regulating signal transduction or  
 CC the expression of a protein capable of acting as a donor or acceptor  
 CC molecule of a phosphate group. The monoclonal antibodies can be used as  
 CC probes for detecting discrete antigens expressed by tissue or cell  
 CC samples, and therefore used in humans for localization and monitoring of  
 CC microbial infection  
 CC

SQ Sequence 5007 BP; 946 A; 1722 C; 1541 G; 798 T; 0 U; 0 Other;

Query Match 93.2%; Score 905.4; DB 5; Length 5007;

Best Local Similarity 99.9%; Pred. No. 1.8e-159;

Matches 906; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	CAGCGGGGGCGCTGGCGCCAGATCATCCCTACCCACCCCAAGCAAGACAGCAGT	60
DB	4101	CAGCGGGGGCGCTGGCGCCAGATCATCCCTACCCACCCCAAGCAAGACAGCAGT	4160
QY	61	GCTCGGGAATACAGAGCCCTCAAGGGCTCGCCACCCCGACCTGCGCCAGCTGCACGC	120
DB	4161	GCTCGGGAATACAGAGCCCTCAAGGGCTCGCCACCCCGACCTGCGCCAGCTGCACGC	4220
QY	121	AGCTTACTTACGCCCCCGGACCTGGTCTCATCTTGGAGCTGTGCTTGGGCGGAGCT	180
DB	4221	AGCTTACTTACGCCCCCGGACCTGGTCTCATCTTGGAGCTGTGCTTGGGCGGAGCT	4280
QY	181	GCTCCCTGCTGCGCGAGAGGGCTCTTACTCAGAAATCCGAGTGAAGGACTACCTGTG	240
DB	4281	GCTCCCTGCTGCGCGAGAGGGCTCTTACTCAGAAATCCGAGTGAAGGACTACCTGTG	4340
QY	241	GCAGATGTTGAGTGGCCACCCAGTACCTGCACACACACATCTCTGCACCTGACCTGAG	300
DB	4341	GCAGATGTTGAGTGGCCACCCAGTACCTGCACACACACATCTCTGCACCTGACCTGAG	4400
QY	301	GTCCGAGAAATGATCATACCGAATACACCTCTCAAGTCTGTGACCTGGGCAATGC	360
DB	4401	GTCCGAGAAATGATCATACCGAATACACCTCTCAAGTCTGTGACCTGGGCAATGC	4460
QY	361	ACAGAGCTTACGACAGAGAGTGTGCTGCTCAGACAAATTCAGAGTACCTAGAGAC	420
DB	4461	ACAGAGCTTACGACAGAGAGTGTGCTGCTCAGACAAATTCAGAGTACCTAGAGAC	4520
QY	421	CATGGCTCCAGAGCTCTGAGAGGCGCCAGGGGCTGTTCCACACACACATCTGGCCAT	480
DB	4521	CATGGCTCCAGAGCTCTGAGAGGCGCCAGGGGCTGTTCCACACACACATCTGGCCAT	4580
QY	481	CGGTGTGACACCTTATCATGTGACGCGCGAGTACCCCGTGAAGAGAGGCTGACG	540
DB	4581	CGGTGTGACACCTTATCATGTGACGCGCGAGTACCCCGTGAAGAGAGGCTGACG	4640
QY	541	CGACCTCAGAGAGGATCGCAGAGGGGCTGTCGCGCTCAGCGCGTGTACCGGGGCT	600
DB	4641	CGACCTCAGAGAGGATCGCAGAGGGGCTGTCGCGCTCAGCGCGTGTACCGGGGCT	4700
QY	601	GTCCGGGGCGCCCGTGCCTTCTTCTGCGCAGACTCTGTGCGCCAGCCCTTGGGCGCGGC	660
DB	4701	GTCCGGGGCGCCCGTGCCTTCTTCTGCGCAGACTCTGTGCGCCAGCCCTTGGGCGCGGC	4760
QY	661	CTGGCGCTCAGCTGCTGAGTCCCGTGTCTTAAACAGAGAGGGCCCGCTGTCGGG	720
DB	4761	CTGGCGCTCAGCTGCTGAGTCCCGTGTCTTAAACAGAGAGGGCCCGCTGTCGGG	4820
QY	721	GCCCGCGCCGCTACCTTCCCTACCGCGCGGCTGCGGCTTCTGTCGCAATCGCGAGAA	780

DB	4821	GCCGCGCGCGGTGACCTTCCCTACCGCGCGGCTGCGGCTTCTCGTGCGCAATCGGAGAA	4880
QY	781	GAGACGCGCGCTGCTGTACAGAGGACACACCTGCGCCAGTGGCTGAGGTTCCGCCCG	840
DB	4881	GAGACGCGCGCTGCTGTACAGAGGACACACCTGCGCCAGTGGCTGAGGTTCCGCCCG	4940
QY	841	GCACACACCTTGGTCTTCCCGCTGGGGGTGCTGTCAGACGCGGCCCAATAAACAACGCCACG	900
DB	4941	GCACACACCTTGGTCTTCCCGCTGGGGGTGCTGTCAGACGCGGCCCAATAAACAACGCCACG	5000
QY	901	CGGGCGA 907	
DB	5001	CGGGCGA 5007	

RESULT 10  
 AAC62287  
 ID AAC62287 standard; cDNA; 7928 BP.  
 XX  
 AC AAC62287;  
 XX  
 DT 19-MAR-2001 (first entry)  
 XX  
 cDNA encoding a splice variant of a signal transduction polypeptide.  
 DE  
 XX  
 KW Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;  
 KW congestive heart failure; dilated congestive cardiomyopathy;  
 KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;  
 KW mitral valve disease; aortic valve disease; tricuspid valve disease;  
 KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;  
 KW atherosclerosis; cardiac tumour; microbial infection; splice variant; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 CDS 60..7850  
 FT /\*tag= a  
 FT /product= "signal transduction polypeptide H19G5 splice  
 FT variant"  
 XX  
 PN WO200063381-A1.  
 XX  
 PD 26-OCT-2000.  
 XX  
 PF 11-APR-2000; 2000WO-US009488.  
 XX  
 PR 16-APR-1999; 99US-0129553P.  
 XX  
 PA (SCIO-) SCTOS INC.  
 XX  
 PI Zeng W, Stanton L, Kong H;  
 DR WPI; 2001-007013/01.  
 DR P-PSDB; AAB30569.  
 XX  
 Novel h19G5 polypeptides capable of regulating signal transduction and  
 PT exhibiting kinase activity useful for identifying antibodies to treat  
 PT cardiac diseases, and additional mediators of signal transduction.  
 XX  
 PS Claim 4; Page 65-68; 81pp; English.  
 CC  
 CC The present sequence encodes a splice variant of human in signal  
 CC transduction polypeptide. The polypeptide is designated H19G5. The  
 CC protein is capable of regulating signal transduction and exhibits kinase  
 CC activity. The H19G5 transcript is expressed in the heart. H19G5  
 CC polypeptides and polynucleotides are useful for preventing or treating a  
 CC cardiac disease, such as congestive heart failure, dilated congestive  
 CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,  
 CC mitral valve disease, aortic valve disease or tricuspid valve disease,  
 CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,  
 CC arterial or renovascular hypertension, arteriosclerosis, atherosclerosis  
 CC and cardiac tumours in humans. The polypeptide is also useful for  
 CC detecting the expression of a protein capable of regulating signal

transduction or the expression of a protein capable of acting as a donor or acceptor molecule of a phosphate group. The monoclonal antibodies can be used as probes for detecting discrete antigens expressed by tissue or cell samples, and therefore used in humans for localization and monitoring of microbial infection

SQ Sequence 7928 BP; 1535 A; 2644 C; 2488 G; 1261 T; 0 U; 0 Other;

Query Match 93.2%; Score 905.4; DB 5; Length 7928;

Best Local Similarity 99.9%; Pred. No. 2e-159;

Matches 906; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	CAGCGGGCGGCGCTGCGCGCAGATATCCCTTACCAACCCGAAGCAAGACAGCAGCT	60
Db	7022	CAGCGGGCGGCGCTGCGCGCAGATATCCCTTACCAACCCGAAGCAAGACAGCAGCT	7081
Qy	61	GCTCGCGGAATACAGAGCCCTCAAGGGCTTGCGCCACCCGACACTGGCCAGCTGCACGC	120
Db	7082	GCTCGCGGAATACAGAGCCCTCAAGGGCTTGCGCCACCCGACACTGGCCAGCTGCACGC	7141
Qy	121	AGCCTACTCAGCCCCCGGCACCTGGTGCTCATCTTGGAGCTGTGCTCTGGGCCCGAGCT	180
Db	7142	AGCCTACTCAGCCCCCGGCACCTGGTGCTCATCTTGGAGCTGTGCTCTGGGCCCGAGCT	7201
Qy	181	GCTCCCTGCTGCTGGCCGAGAGGGGCTCTACTCTAGAACTCCGAGGTGAAGACTACTCTGTG	240
Db	7202	GCTCCCTGCTGCTGGCCGAGAGGGGCTCTACTCTAGAACTCCGAGGTGAAGACTACTCTGTG	7261
Qy	241	GCAGATGTGAGTGCCACCCAGTACTTGCAAAACAGACATCTCTGCACCTGGACCTTGAG	300
Db	7262	GCAGATGTGAGTGCCACCCAGTACTTGCAAAACAGACATCTCTGCACCTGGACCTTGAG	7321
Qy	301	GTCGGAACATGATCATCACCGAATAACAACCTGCTCAAGTGTGTGACCTGGGCAATGC	360
Db	7322	GTCGGAACATGATCATCACCGAATAACAACCTGCTCAAGTGTGTGACCTGGGCAATGC	7381
Qy	361	ACAGAGCCTCAGCCAGGAGAAAGTGCTGCCCTCAGACAAGTTCAAGGACTACTTAGAGAC	420
Db	7382	ACAGAGCCTCAGCCAGGAGAAAGTGCTGCCCTCAGACAAGTTCAAGGACTACTTAGAGAC	7441
Qy	421	CATGGCTCCAGAGCTCCTCGAGGGCCAGGGGCTGTCCACAGACAGACATCTGGGCCAT	480
Db	7442	CATGGCTCCAGAGCTCCTCGAGGGCCAGGGGCTGTCCACAGACAGACATCTGGGCCAT	7501
Qy	481	CGGTGTGACAGCCTTCATCATGCTGAGCGCCGAGTACCCGGTGAGCAGCGGGTGCACG	540
Db	7502	CGGTGTGACAGCCTTCATCATGCTGAGCGCCGAGTACCCGGTGAGCAGCGGGTGCACG	7561
Qy	541	CGACTGACAGAGAGACTGCGCCAGAGGGCTGGTCCGGCTGAGCCGCTGCTACGCGGGCT	600
Db	7562	CGACTGACAGAGAGACTGCGCCAGAGGGCTGGTCCGGCTGAGCCGCTGCTACGCGGGCT	7621
Qy	601	GTCGCGGGGCGCGTGCGCTTCTTCCGCGAGACACTGTGTGCGCCAGCCCTGGGGCCGCGCC	660
Db	7622	GTCGCGGGGCGCGTGCGCTTCTTCCGCGAGACACTGTGTGCGCCAGCCCTGGGGCCGCGCC	7681
Qy	661	CTGCGGCTCAGCTGCTGTGAGTGCCCGTGCGTAAACAGAGAGGGCCCGGCTGTTCGCG	720
Db	7682	CTGCGGCTCAGCTGCTGTGAGTGCCCGTGCGTAAACAGAGAGGGCCCGGCTGTTCGCG	7741
Qy	721	GCCCGCGCCGTGACTTTCCTTACCGCGCGCTGCGCGCTCTTCTGTGCGCAATCCGAGAA	780
Db	7742	GCCCGCGCCGTGACTTTCCTTACCGCGCGCTGCGCGCTCTTCTGTGCGCAATCCGAGAA	7801
Qy	781	GAGACGCGCGTGTGTAAAGAGCAAACTGGGCCCAGGTGCGCTGAGGGTGCCTCCCGC	840
Db	7802	GAGACGCGCGTGTGTAAAGAGCAAACTGGGCCCAGGTGCGCTGAGGGTGCCTCCCGC	7861
Qy	841	GCCACACCTTGGTCTCCCGCTGGGGTGTGCTGCGAGACGCGCCAAATAAAACGCCACGC	900
Db	7862	GCCACACCTTGGTCTCCCGCTGGGGTGTGCTGCGAGACGCGCCAAATAAAACGCCACGC	7921
Qy	901	CGGGCGA	907

RESULT 11

RESULT 11  
ABN76255  
ID ID ABN76255 standard; cDNA; 225 BP.  
XX XX  
AC AC ABN76255;  
XX XX  
DT DT 08-JUL-2002 (first entry)  
XX XX  
DE DE Human kinase-like ORF1202 cDNA, SEQ ID NO:2403.  
XX XX  
KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
KW immune modulation; haematopoiesis regulation; tissue growth;  
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; tumour inhibition; bodily characteristics; fertility;  
KW behaviour; cancer; proliferative disorder; neurological disorder;  
KW cardiovascular disease; immune system disorder; organ transplantation;  
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
KW vasotrophic; antipsoriatic; antidiabetic; cytostatic; neurotropic;  
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;  
KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.  
XX XX  
OS OS Homo sapiens.  
XX XX  
FN WO200190366-A2.  
XX XX  
PD 29-NOV-2001.  
XX XX  
PF 24-MAY-2001; 2001WO-US017076.  
XX XX  
PR 24-MAY-2000; 2000US-020690P.  
XX XX  
FA (CURA-) CURAGEN CORP.  
XX XX  
FI Leach MD, Shinkets RA;  
XX XX  
DR WPI; 2002-106200/14.  
DR P-PSDB; ABP32229.  
XX XX  
PT Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and disorders related to organ  
PT transplantation.  
XX XX  
PS Claim 1; Page 863; 2508pp; English.  
XX XX  
CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
CC ABN79587 represent cDNAs encoding them. The invention also encompasses  
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
CC polypeptides, methods of screening for modulators of ORFX expression or  
CC activity and methods of screening individuals for a predisposition to an  
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
CC range of biological activities, such as cytokine, cell proliferation,  
CC cell differentiation, immune modulation, haematopoiesis regulation,  
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
CC and antiinfective activity, and may also be involved in the determination  
CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
CC nucleic acids and antibodies may be used in the treatment of cancers,  
CC neuroproliferative disorders such as psoriasis and benign tumours,  
CC other proliferative disorders such as epilepsy and Alzheimer's disease,



CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes, in the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases  
 XX  
 SQ Sequence 225 BP; 54 A; 65 C; 69 G; 37 T; 0 U; 0 Other;

Query Match 23.0%; Score 223.4; DB 6; Length 225;

Best Local Similarity 99.6%; Pred. No. 1.4e-32;  
 Matches 224; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 314 ATCATCACCAGTAACAACCTGCTCAAGTCTGTCGACCTGGGCAATGCACAGAGCTCAGC 373  
 DB |||||  
 QY 374 CAGGAGAGGTGCTGCTCCTCAGACAAAGTTCAAGACTACCTAGACACATGGCTCCAGAG 433  
 DB 61 CAGGAGAGGTGCTGCTCCTCAGACAAAGTTCAAGACTACCTAGACACATGGCTCCAGAG 120  
 QY 434 CTCCTGGAGGGCCAGGGGGCTGTTCCACAGACACATCTGGGGCATCGGTGTGACAGCC 493  
 DB 121 CTCCTGGAGGGCCAGGGGGCTGTTCCACAGACACATCTGGGGCATCGGTGTGACAGCC 180  
 QY 494 TTATCATGCTGAGCGCCAGTACCCGGTGAGCAGCGAGGTGCA 538  
 DB 181 TTCATCATGCTGAGCGCCAGTACCCGGTGAGCAGCGAGGTGCA 225

## RESULT 12

AAAF4662

ID AAFA4662 standard; cDNA; 7710 BP.

AC AAFA4662;

DT 27-MAR-2001 (first entry)

DE Novel protein kinase cDNA, SEQ ID NO: 42.

XX Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;  
 KW immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;  
 KW dermatological; antidiabetic; antifertility; gene therapy; vaccine;  
 KW immune disorder; cardiovascular disease; neurodegenerative disease;  
 KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;  
 KW inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.

OS Homo sapiens.

PN WO2000073469-A2.

PD 07-DEC-2000.

XX 26-MAY-2000; 2000WO-US014842.

XX 28-MAY-1999; 99US-0136503P.

XX (SUGEN-) SUGEN INC.

XX Plowman GD, Martinez R, Whyte D, Sudersanam S;

XX WPI; 2001-032161/04.

DR P-PSDB; AAB65635.

XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and  
 PT treating immune-related diseases and disorders, cardiovascular disease,

PT neurodegenerative diseases and/or cancers.

XX Example 1; Fig 2; 310pp; English.

XX The present sequence encodes a novel protein kinase. The nucleic acids  
 CC and the protein kinases they encode may be used in the treatment and  
 CC diagnosis of diseases associated with inappropriate kinase expression  
 CC such as immune-related diseases and disorders, cardiovascular disease,  
 CC neurodegenerative diseases and/or cancers. The nucleic acids and  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC assays. The kinase polypeptides may be used as antigens in the production  
 CC of antibodies of kinase expression and activity. Anti-kinase antibodies  
 CC and kinase antagonists may also be used to down regulate kinase  
 CC expression and activity. Diseases related to kinase expression and  
 CC activity include rheumatoid arthritis, atherosclerosis, autoimmune  
 CC disorders, complications of organ transplantation, myocardial infarction,  
 CC immune disorders, cardiomyopathies, strokes, renal failure, oxidative-  
 CC stress related disorders, chronic inflammatory bowel disease, chronic  
 CC inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis,  
 CC psoriasis, rhinitis, autoimmunity, diabetes, cancers and reproductive  
 CC disorders  
 XX

SQ Sequence 7710 BP; 1446 A; 2534 C; 2447 G; 1283 T; 0 U; 0 Other;

Query Match 21.2%; Score 206; DB 4; Length 7710;

Best Local Similarity 53.3%; Pred. No. 4e-29;

Matches 434; Conservative 0; Mismatches 380; Indels 0; Gaps 0;

QY 1 CAGCGGGGGGGCGGCGCCCAAGATCATCCCTTACCACCCCAAGGACACAGACGAGT 60  
 DB 6018 CAGCGGGCGAAGCTTCGTGGGCCAAGATCGTCCCTATCTGCCAGGCAAGCCCGGGT 6077  
 QY 61 GTTCGCGGAATACGAGGGCCCTCAAGGGCTCGGCCACCCGACCTGGCCCAAGCTGCACG 120  
 DB 6078 CCGCAGGAGTACGAGGTGCTGGGAGCCCTGCACACGAGCGGATCATGTCCCTGCACGA 6137  
 QY 121 AGCTTACTCAGCCCCCGGACCTGGTCTCATCTTGGAGCTGTCTTGGGCCCGAGCT 180  
 DB 6138 GGCCTACATACCCCTCGGTACCTCGTGTCTCATCTTGGAGCTGTGGCAACCGGGAAT 6197  
 QY 181 GCTCCCTCGCTGGCGGAGAGGGCTCCTCTACTCAAGATCCAGGTGAGGACTACCTGTG 240  
 DB 6198 CCGTGTGGGCTCAGTGACAGGTTCCGGTATTCTGAGGATGACGTGGCCACTTACATGT 6257  
 QY 241 GCAGATGTTGAGTGCCACCCAGTACCTGCACAAACAGACATCTCTGCACCTGGACCTGAG 300  
 DB 6258 GCAGTGTCTAAGGCTGGACTACCTCCAGGGCCACCCAGCTGTCCACCTAGACATCAA 6317  
 QY 301 GTCGAGAACATGATCATCACCAGTAACAACCTGTCAAGTCTGTGACCTGGGCAATGC 360  
 DB 6318 GCCAGAACCTGTGCTGGCCCTGACAAATGCCCTCAAGATTGTGACCTTGGCAGTGC 6377  
 QY 361 ACAGAGCTCTCAGCCAGGAGAGGTGCTGCTCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 420  
 DB 6378 CCAGCCCTACAAACCCCGGAGGCTTAGGCCCCCTTGGCCACCGCACGCGCTGGAGTT 6437  
 QY 421 CATGGCTCCAGAGCTCCTGGAGGGCCAGGGGGCTGTTCCACAGACACATCTCGGCAAT 480  
 DB 6438 CATGGCTCCGAGATGGTGAAGGAGAACCCATCGGCTCTGCCACGACATCTGGGAGC 6497  
 QY 481 CGGTGTGACAGCTTCATCATCTGAGCGCCGAGTACCCGGTGAGCAGCGAGGTGCAACG 540  
 DB 6498 GGGTGTGCTCACTTACATTATGCTCAGTGAGCGCTGCCCGTCTCTATGAGCAGACCCCCA 6557  
 QY 541 CGACCTGCAGAGAGACTGCCCAAGGGGCTGGTCCGGCTGAGCGCTGTCTACCGGGGCT 600  
 DB 6558 GGAACGAGAGGCTCGGATTTGGGGGCGCGCTTTGATGCTTCCAGTGTACCCCAATAC 6617  
 QY 601 GTCCGGGGGCGCGGTGGCCCTTCTGCGCAGCACTCTGTGCGCCAGCCCTTGGGGCGCGCC 660  
 DB 6618 ATCCAGAGCGCCACCCCTCTCTTTCGGAAGGTTCTCTCTGTACATCCCTGGAGCGCGCC 6677  
 QY 661 CTGCGCTGCCAGTGCCTGTGAGTCCCGTGGCTAACAGAGAGGGCGCGCGCTTTCGCG 720



Db 6678 CTCCTGTCAGGACGCTGCGCCACCATGTTTCAGAGCGCTTACCTGATGAAGTGGC 6737  
Qy 721 GCCCGCGCGCTGACCTTCCCTACCGCGCGGCTCGGGTCTTCGTGGCGCAATCGCGAGAA 780  
Db 6738 CCGCAGACGCTCACCTTCACCAACCGGCTCAAGAGTTCCTGGCGAGCGCGG 6797  
Qy 781 GAGACGCGCGTGTGTACAGAGGACACACTG 814  
Db 6798 GCGCGCGGCTGAGGCTGCCACCGCCACCAAGGTG 6831

RESULT 13

AAD30565 standard; cDNA; 7789 BP.

XX AAD30565;  
XX 21-MAY-2002 (first entry)  
XX Human kinase polypeptide (PKIN-18) cDNA.  
XX Human; kinase polypeptide; PKIN-18; gene therapy; Addison's disease;  
KW leukaemia; immune disorder; lymphoma; melanoma; developmental disorder;  
KW acquired immunodeficiency syndrome; AIDS; allergy; anaemia; hypertension;  
KW asthma; Crohn's disease; rheumatoid arthritis; bursitis; atherosclerosis;  
KW cirrhosis; hepatitis; psoriasis; Cushing's syndrome; cytostatic; cancer;  
KW cholestasis; cardiac; cardiovascular disorder; Niemann-Pick's disease;  
KW lipid disorder; fatty liver; Gaucher's disease; myocardial infarction;  
KW drug screening; transgenic animal; antiinflammatory; hepatotropic;  
KW hypotensive; anti-HIV; enzyme; ss.

OS Homo sapiens.

XX WO200208399-A2.

XX 31-JAN-2002.

XX 20-JUL-2001; 2001WO-US023092.

XX 21-JUL-2000; 2000US-0220038P.

XX 28-JUL-2000; 2000US-0222112P.

XX 04-AUG-2000; 2000US-0222831P.

XX 11-AUG-2000; 2000US-0224729P.

XX (INCY-) INCYTE GENOMICS INC.

PA (THOR/) THORNTON M.

XX Thornton M, Yue H, Khan FA, Gururajan R, Hafalia AJA, Wallia NK;  
PI Patterson C, Rankumar J, Gandhi AR, Policky JL, Baughn MR;  
PI Tribouley CM, Bandman O, Nguyen DB, Lu Y, Burford N, Lal P, Ding L;  
PI Yao MG, Elliott VS, Recipon SA, Kearney L, Lu DAM, Greenwald SR;  
PI Tang YT, Xu Y, Walsh RT, Gietzen KJ, Yang J, Hillman JL;

XX WPI; 2002-206083/26.

XX New human kinase polypeptide, useful in diagnosis, prevention and  
PT treatment of cancer, immune disorder, growth and developmental disorder,  
PT cardiovascular disorder and lipid disorder.

XX Claim 5; Page 191-193; 196pp; English.

XX The present invention relates to an isolated human kinase polypeptide  
CC (PKIN) or a biologically active/immunogenic fragment of PKIN. PKIN is  
CC useful for diagnosing, treating and preventing cancer (e.g., leukaemia,  
CC lymphoma, melanoma), an immune disorder (e.g., acquired immunodeficiency  
CC syndrome (AIDS), Addison's disease, allergy, anaemia, asthma, Crohn's  
CC disease, rheumatoid arthritis), a growth and developmental disorder (e.g.  
CC bursitis, cirrhosis, hepatitis, psoriasis, Cushing's syndrome), a  
CC cardiovascular disorder (e.g., atherosclerosis, hypertension, myocardial  
CC infarction), and a lipid disorder (e.g., fatty liver, cholestasis,  
CC Gaucher's disease, Niemann-Pick's disease). PKIN is useful in a number of  
CC drug screening techniques and to analyse the proteome of a tissue or cell

CC type. PKIN is useful for creating knockin humanised animals or transgenic  
CC animals to model human diseases, in somatic or germline gene therapy, to  
CC generate a transcript image of a tissue or cell type, for detecting  
CC differences in the chromosomal location due to translocation, inversion,  
CC etc., among normal, carrier or affected individuals, and as hybridisation  
CC probes for mapping naturally occurring genomic sequences. PKIN is useful  
CC in southern or northern analysis, dot blot or other membrane-based  
CC technologies, in PCR technologies, in dipstick, pin, multiformat enzyme  
CC linked immunosorbent (ELISA)-like assays and in microarrays utilising  
CC fluids or tissues from patients to detect altered PKIN expression. The  
CC present sequence is human PKIN-18 cDNA. Note: This sequence is said to  
CC encode PKIN-18 referred as SEQ ID NO:18 (AAE19160). However this does not  
CC appear to be the case

XX Sequence 7789 BP; 1459 A; 2558 C; 2470 G; 1292 T; 0 U; 0 Other;  
SQ

Query Match 21.1%; Score 204.4; DB 6; Length 7789;  
Best Local Similarity 53.2%; Pred. No. 8e-29;  
Matches 433; Conservative 0; Mismatches 381; Indels 0; Gaps 0;

Qy 1 CAGCGGGCGGCGCTGGCGCAAGATCATCCCTACCACCCACAGGACAAAGACGACGT 60  
Db 6080 CACGGGGCAAGATTCTGTCGGCCAGATCTGTGCTATGTGTCGAGGGCAAGCGGGGT 6139  
Qy 61 GTCCGCGCAATACGAGGCGCTCAAGGGCTGCGCCACCGACCTGGCCAGCTGCACGC 120  
Db 6140 CTTGCAGGAGTACGAGGTGCTCGGACCTGCACACGAGCGGATCATGTCCCTGCACGA 6199  
Qy 121 AGCTACCTCAGCCCGGCGACCTGCTGCTCATCTTGGAGCTGTGCTGGCCCGAGCT 180  
Db 6200 GGCTCATACACCCCTCGGTACTCTGCTCATCTGCTGAGAGCTGTGGCAACCGGAAC 6259  
Qy 181 GCTCCCTGCTGGCGGAGAGGCGCTCTTACTCAGAAATCCGAGGTCGAAGACTTACCTGTG 240  
Db 6260 CCTCTGCGGCTCAGTGACAGGTTCCGGTATTCTGAGGATGACGTGCCACATACATGGT 6319  
Qy 241 GCAGATGTTGAGTGCCACCCAGTACTGCACAAACAGACATCTCTGACCTGACCTGAG 300  
Db 6320 GCAGCTGCTACAGGCTCGACTTACCTCCACGCGCCACACGCTCCACCTAGACATCAA 6379  
Qy 301 GTCGAGAGACATGATCATCACGAAATACAACTGCTCAAGGTCTGTCGAGCTGGGCAATGC 360  
Db 6380 GCAGACACCTGCTGCTGGCCCTGACAAATGCCCTCAGATGTGAGACTTTGGCAGTGC 6439  
Qy 361 ACAGAGCCTCAGCCAGGAGAGGTGTCCTCAGACAAAGTTCAAGGACTACTTAGAGAC 420  
Db 6440 CCAAGCCTACACCCCGCGGCGCTTGGCGCCCTTGGCCACCGGCGGACGCTGGAGTT 6499  
Qy 421 CATGGCTCCAGAGCTCTCGAGGCGCCAGGGGGTGTTCACAGACAGACATCTGGGCCAT 480  
Db 6500 CATGGCTCCGAGATGTTGAGGGAGAACCCATCGGCTCTGCCACGAGACATCTGGGGAGC 5559  
Qy 481 CGGTGTACAGCCTTTCATCATGTGAGCGCCGAGTACCCCGGTGAGCAGCGAGGTCACG 540  
Db 6560 GGGTGTGCTCACCTTACATATGCTCAGTGGACGCTCCCGTTCTATGAGCCAGACCCCA 6619  
Qy 541 CGACCTGCAGAGGAGTCCGCAAGGGGCTGGTCCGGCTGAGCGGCTGTACCGGGGCT 600  
Db 6620 GGAACCGAGGCTCGGATTTGTGGGGGGCGGCTTGTATGCTTCAGCTGTACCCCAATAC 6679  
Qy 601 GTCGCGGGGCGCGTGGCGCTTCTCGCGAGCACTCTGTGCGCCCGACGCTTGGGGCGCGGC 660  
Db 6680 ATCCAGAGCGCCACCTCTTCTTGCAGAGGTTCTCTGTATCATCTCTGGAGCGCGGC 6739  
Qy 661 CTGCGCGTCCAGCTGCTGAGTCCCGTGGCTGTAACAGAGGAGGCGCGGCTGTTCGCG 720  
Db 6740 CTCCTCTGAGGACTGCTGCTGGCCACCCATGTTTGCAGGACGCTTACTGATGAAGTGG 6799  
Qy 721 GCCCGCGCGCTGACCTTCCCTACCGGGGCTGCGGCTTCTGTCGGAATCGCGAGAA 780  
Db 6800 CCGCCAGACCTCACCTTACACCAACCGGCTCAGAGAGTTCTTGGCGGAGCAGCGCGG 6859  
Qy 781 GAGACGCGCGCTGCTGTACAGAGGACACAACTG 814



KW diabetes mellitus; viral; bacterial; fungal; parasitic; protozoan;  
KW helminthic infection; transgenic; gene therapy; human; ss; gene.  
XX Homo sapiens.  
XX WO2003033680-A2.  
XX 24-APR-2003.  
XX 17-OCT-2002; 2002WO-US033723.  
XX 19-OCT-2001; 2001US-0345474P.  
XX 02-NOV-2001; 2001US-0343910P.  
XX 13-NOV-2001; 2001US-0333098P.  
XX 16-NOV-2001; 2001US-0332424P.  
XX 30-NOV-2001; 2001US-0334288P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX Bandman O, Baughn MR, Becha SD, Borowsky ML, Duggan BM;  
XX Emerling BM, Forsythe IJ, Gandhi AR, Gorvad AE, Griffin JA;  
XX Gururajan R, Hafalia AJA, Khan FA, Lal PG, Lee EA, Lee SY;  
XX Lindquist EA, Lu DAM, Lu Y, Marquis JP, Nguyen DB, Arvizu CS;  
XX Ramkumar J, Recipon SA, Richardson TW, Swarnakar A, Tang YT;  
XX Thornton MB, Tran UK, Chawla NK, Warren BA, Yang J, Yao MG, Yue H;  
XX Zebarjadian Y;  
XX WPI; 2003-403214/38.  
XX P-PSDB; ADC99075.  
XX New human kinases and phosphatases and polynucleotides, useful for  
XX diagnosing, treating or preventing autoimmune or inflammatory disorders  
XX (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,  
XX cancer or hepatitis.  
XX Claim 5; SEQ ID NO 80; 424pp; English.  
XX The invention relates to a novel isolated polypeptide which is a human  
XX kinase and phosphatase (KPP). The KPP polypeptides, polynucleotides,  
XX agonists and antagonists are useful for diagnosing, treating or  
XX preventing cell proliferative disorders such as atherosclerosis,  
XX cirrhosis, hepatitis and cancer, developmental disorders e.g. mental  
XX retardation, neurological disorders including Alzheimer's disease and  
XX Parkinson's disease, autoimmune and inflammatory disorders such as  
XX Crohn's disease and diabetes mellitus and finally, viral, bacterial, the  
XX fungal, parasitic, protozoan or helminthic infections. Furthermore, the  
XX polynucleotides encoding KPP may be useful for creating transgenic  
XX animals to model human disease, as well as during gene therapy  
XX procedures. The current sequence is that of the human KPP cDNA of the  
XX invention.

SQ Sequence 10662 BP; 1933 A; 3587 C; 3486 G; 1656 T; 0 U; 0 Other;

Query Match 21.1%; Score 204.4; DB 9; Length 10662;  
Best Local Similarity 53.2%; Pred. No. 8.4e-29;  
Matches 433; Conservative 0; Mismatches 381; Indels 0; Gaps 0;

QY 1 CAGCGGGGGGGCTGGCGCCAGATCATCCCTACCAACCCCAAGGACACACAGCAGT 60  
DB 8961 CAGCGGGGGAACCTTCGTGGCCAAAGATCGTCCCTATGTGCGGAGGGGACGGCGGGT 9020  
QY 61 GCTGCGCGAATACAGCGCCCTCAAGGGGCTGGCGCCACCCCGACCTGGCCACGTCACGC 120  
DB 9021 CTTGCCAGGATACAGAGTGTGGGACCTGTCACACAGCGGATCATGTCCCTGCACGA 9080  
QY 121 AGCCTACCTCAGCCCGGACCTGTGCTCATCTTGGAGCTGTGCTCTGGGCGCGAGCT 180  
DB 9081 GGCTTACATCACCCCTCGGTACCTCGTGTCTATTGTGTGAGAGCTGTGGCAACCGGGAAT 9140  
QY 181 GCTCCCTCGCTGGCGAGAGGGCCCTCCTACTCAGAAATCCGAGGTGAAGGACTACTGTG 240  
DB 9141 CCTCTGTGGCTCAGTGACAGGTTCGGGTATCTTGAGGATGCTGGCCACTTACATGCT 9200

QY 241 GCAGATGTTAGTGGCCACCCAGTACTCTGCAAAACGAGACATCTTGACCTGACCTGAG 300  
DB 9201 GCAGCTGTCTACAAGGCTTGGACTACTCTCCACGGCCACCAACGCTGCTCCACCTAGCA 9260  
QY 301 GTCCGAGACATGATCATCACCGAATAACAACCTGCTCAAGGTCGTGGACCTGGCAATGC 360  
DB 9261 GCCAGACAACCTGCTGCTGGCCCTCGACAAATGCCCTCAAGATTGTGGACTTTGGCAGTGC 9320  
QY 361 ACAGAGCCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACTTAGAGAC 420  
DB 9321 CCAGCCCTACAACCCCGAGGCTTGGGCCCTTGGCCACCGCAGGGGACGCTGGAGTT 9380  
QY 421 CATGGCTCCAGAGCTCCTGGAGGGCCAGGGGCTGTTCACAGACAGACATCTGGGCCAT 480  
DB 9381 CATGGCTCCGGAGATGGTGAAGGGAGAACCCATTCGGCTCTGCCACGACATCTGGGGAGC 9440  
QY 481 CGGTGTGACAGCCTTCATCATGCTGAGCGCCGAGTACCCCGGTGAGCAGCGAGGGTGCAG 540  
DB 9441 GGGTGTGCTCACTTACATTTATGCTCAGTGGACGCTCCCGCTTCTATAGCCAGACCCCA 9500  
QY 541 CGACCTGCGAGAGGACTCGGCAAGGGGTGTTCGGCTGAGCGCGTGTCTAGCGGGGCT 600  
DB 9501 GGAACCGAGGCTCGGATTGTGGGGGGCGCTTGTGATGCTTCCAGCTGTACCCCAATAC 9560  
QY 601 GTCCGGGGGGCGCGTGGCCTTCTGCGCAGCACTCTGTGCGCCCAAGCCCTGGGGCGGGCC 660  
DB 9561 ATCCAGAGCGCCACCTCTTCTTGGAAAGTTCTCTCTGTACATCCCTGGAGCGGGCC 9620  
QY 661 CTCGGCTCAGCTGCTGCTGAGTGGCCGTGCTTAACAGAGAGGGGCGCGCCTGTTCGG 720  
DB 9621 CTCCTGCAAGACTGCTGCTGCGCCACCCATGGTTGCAAGACGCTTACTGATGAAGTGG 9680  
QY 721 GCCCGGCGCGTGCACCTTCCCTACCGGGGCTGCGCGTCTTCGTGCGCAATCGCGAGAA 780  
DB 9681 CCGCCAGACCTCACCTTACACCAACCGGCTCAAGGAGTTCTCTGGGGGAGCAGCGGCG 9740  
QY 781 GAGACGCGCGCTGCTGTATCAAGAGGACCAACCTG 814  
DB 9741 GCGCGGGCTGAGGCTGCGCCACCGCCACCAAGGTG 9774

Search completed: September 19, 2004, 15:42:26  
Job time : 404.361 secs

**This Page Blank (uspto)**

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 12:53:28 ; Search time 158.742 Seconds  
(without alignments)  
10303.209 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_16862\_17246

Perfect score: 385  
Sequence: 1 ccgaggagagatctttgaca.....acctgcagcaccctggagcgc 385

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseqn\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	385	100.0	7928	5 AAC62287	AAC62287 cDNA enco
2	385	100.0	8106	7 ABX11641	Abx11641 Human ser
3	385	100.0	24120	7 ABX11642	Abx11642 Human ser
4	377	97.9	628	3 AAC74767	AAC74767 Human ORF
5	377	97.9	628	6 AEN75689	AEN75689 Human ORF
6	42.6	11.1	4180	6 ABK84118	ABK84118 Human cDN
7	42.6	11.1	4282	6 ABS51830	ABS51830 Novel hum
8	42.6	11.1	4366	7 ABX63668	ABX63668 Human cDN
9	41.4	10.8	1803	5 AAC84312	AAC84312 Human EXC
10	39.8	10.3	1278	7 ACA23561	ACA23561 Prokaryot
11	39.4	10.2	455	8 ACH34588	ACH34588 Human end
12	39.4	10.2	474	8 ACH34675	ACH34675 Human end
13	39.4	10.2	1302	6 ABS55941	ABS55941 DNA topoi
14	39.4	10.2	1354	4 AAS26907	AAS26907 Human cDN
15	39.4	10.2	1354	6 ABQ54643	ABQ54643 Human ova
16	39.4	10.2	1371	4 AAS26834	AAS26834 Human cDN
17	39.4	10.2	1720	7 ABX05065	ABX05065 Human nov
18	39.4	10.2	2040	9 ADB63018	ADB63018 Human cDN
19	39.4	10.2	2151	5 ABA15210	ABA15210 Human ner
20	39.4	10.2	2156	5 ABA15209	ABA15209 Human ner
21	39.4	10.2	125401	4 AAD17186	AA17186 Streptomy
22	38.8	10.1	1131	7 ACA25328	ACA25328 Prokaryot
23	38.6	10.0	307	6 ABL79596	ABL79596 Human Ova

24	38.6	10.0	21185	3 AAA63350	AAA63350 Streptomy
25	38.6	10.0	63164	3 AAA63348	AAA63348 Streptomy
26	38	9.9	1032	7 ADA69587	ADA69587 Rice gene
27	37.8	9.8	5154	5 AAS84859	AAS84859 DNA enco
28	37.8	9.8	5751	4 AAK51986	AAK51986 Human pol
29	37.6	9.8	3525	4 ABL09775	ABL09775 Drosophil
30	37.6	9.8	8848	4 ABL09774	ABL09774 Drosophil
31	37.2	9.7	1200	3 AAA30024	AAA30024 Human PRO
32	37.2	9.7	1200	3 AAA77525	AAA77525 Human PRO
33	37.2	9.7	1200	3 ADC78342	ADC78342 Human PRO
34	37.2	9.7	1200	4 AAF72376	AAF72376 Human PRO
35	37.2	9.7	1200	4 AAC87033	AAC87033 Nucleotid
36	37.2	9.7	1200	4 AAS21380	AAS21380 Human cDN
37	37.2	9.7	1200	4 AAF58494	AAF58494 PRO187 co
38	37.2	9.7	1200	7 ACA58913	ACA58913 Human PRO
39	37.2	9.7	1200	7 ACD23989	ACD23989 Novel hum
40	37.2	9.7	1200	7 ACA58310	ACA58310 cDNA enco
41	37.2	9.7	1200	7 ACD42383	ACD42383 Novel hum
42	37.2	9.7	1200	7 ACA67130	ACA67130 cDNA enco
43	37.2	9.7	1200	7 ACA60017	ACA60017 Human cDN
44	37.2	9.7	1200	7 ACD07417	ACD07417 Novel hum
45	37.2	9.7	1200	7 ACA03739	ACA03739 cDNA enco

#### ALIGNMENTS

RESULT 1  
AAC62287

ID AAC62287 standard; cDNA; 7928 BP.

AC AAC62287;

DT 19-MAR-2001 (first entry)

XX cDNA encoding a splice variant of a signal transduction polypeptide.

XX Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;  
congestive heart failure; dilated congestive cardiomyopathy;  
hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;  
mitral valve disease; aortic valve disease; tricuspid valve disease;  
myocardial infarction; cardiac arrhythmia; arteriosclerosis;  
atherosclerosis; cardiac tumour; microbial infection; splice variant; ss.  
Homo sapiens.

XX Key Location/Qualifiers  
XX CDS 60..7850

XX FT /\*tag= a  
XX FT /product= "signal transduction polypeptide H19G5 splice variant"

XX WO2000063381-A1.

XX PD 26-OCT-2000.

XX PF 11-APR-2000; 2000WO-US009488.

XX PR 16-APR-1999; 99US-0129553P.

XX PA (SCIO-) SCIOS INC.

XX PI Zeng W, Stanton L, Kong H;

XX P-PSDB; 2001-007013/01.

XX DR AAB30569.

XX PT Novel h19G5 polypeptides capable of regulating signal transduction and  
exhibiting kinase activity useful for identifying antibodies to treat  
cardiac diseases, and additional mediators of signal transduction.

XX PS Claim 4; Page 65-68; 81pp; English.



Db 908 GCAGGATGCCATCAGCTGCGGAGGCCAGCTATGTGGAGTCTGTGATGAGCCACCC 967  
QY 121 ACTGCGCTGGCTTGTCCCGCACCAAGCCCAAGTCCAGCCCTCACGCGAGGGCTGGT 180  
Db 968 ACTGCGCTGGCTTGTCCCGCACCAAGCCCAAGTCCAGCCCTCACGCGAGGGCTGGT 1027  
QY 181 GTCACAGCTTACCTGAGCAGAGGCTCAAGTGTCACTGAGTGGGGGGCGCTGAGGC 240  
Db 1028 GTCACAGCTTACCTGAGCAGAGGCTCAAGTGTCACTGAGTGGGGGGCGCTGAGGC 1087  
QY 241 CCCTGAGTTCCTGGGAGGCTGTCTGAGAGCGATACAGCAAGGCTGAGCTCTGT 300  
Db 1088 CCCTGAGTTCCTGGGAGGCTGTCTGAGAGCGATACAGCAAGGCTGAGCTCTGT 1147  
QY 301 GATCCAGAGCTGTGAGTCTTGAGCAGGCCCTTCGTGGAGGAGCTGAGTTCCTGCGAG 360  
Db 1148 GATCCAGAGCTGTGAGTCTTGAGCAGGCCCTTCGTGGAGGAGCTGAGTTCCTGCGAG 1207  
QY 361 CCACCACTGAGCAGCTGGAGCGC 385  
Db 1208 CCACCACTGAGCAGCTGGAGCGC 1232

RESULT 3  
ABX11642  
ID ABX11642 standard; cDNA; 24120 BP.  
XX  
AC ABX11642;  
XX  
DT 09-MAY-2003 (first entry)  
XX  
DE Human serine/threonine or protein kinase 12599, cDNA.  
XX  
KW Human; ss; gene; serine/threonine kinase; protein kinase; 12599;  
KW cardiovascular disease; heart failure; myocardial infarction;  
KW blood vessel disorder; atherosclerosis; Kaposi's sarcoma;  
KW blood platelet disorder; thrombocytopaenia; leukaemia; Hodgkin's disease;  
KW haemolytic anaemia; cellular proliferative disorder; cancer;  
KW protein kinase disorder; autoimmune disorder; diabetes mellitus;  
KW psoriasis; inflammatory bowel disease; rheumatoid arthritis;  
KW multiple sclerosis.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FH 5'UTR 1..71  
FT Location/Qualifiers  
FT /tag= a  
FT 72..23978  
FT /tag= b  
FT /product= "Kinase 12599"  
FT /note= "This CDS is specifically claimed in claim 2"  
FT 23979..24120  
FT /tag= c  
XX  
3'UTR  
XX  
US2002168742-A1.  
XX  
PD 14-NOV-2002.  
XX  
PF 15-FEB-2002; 2002US-00077130.  
XX  
PR 15-FEB-2001; 2001US-0269201P.  
XX  
FA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Kapeller-Libermann R, Acton SL;  
XX  
DR WPI; 2003-298729/29.  
DR P-FSDB; ABG76187.  
XX  
PT Novel isolated human protein kinase, designated 59079 or 12599  
PT polypeptide, useful as diagnostic and therapeutic agents for preventing  
PT cardiovascular diseases, proliferative disorders, and protein kinase

disorders.  
XX  
PS Claim 2; Page 58-84; 119pp; English.  
XX  
CC The invention relates to an isolated human serine/threonine or protein  
CC kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule  
CC comprising at least 85% identity to the nucleic acids appearing as  
CC ABX11641 and ABX11642 or their complement, a naturally occurring variant  
CC of the kinases or their fragments. Also included are a non-human host  
CC cell containing the nucleic acids, an antibody specific for the proteins,  
CC identifying a compound which binds to the kinase (by contacting the  
CC kinase or a cell expressing the kinase with a test compound and  
CC determining whether the kinase binds to the test compound) and modulating  
CC the activity of kinase using the identified compound. The kinases and  
CC their encoding nucleic acids are useful as diagnostic and therapeutic  
CC agents for preventing a disease or condition associated with an aberrant  
CC or unwanted 59079 or 12599 activity in a subject, including  
CC cardiovascular diseases such as heart failure, and myocardial infarction;  
CC disorders involving blood vessels such as atherosclerosis, and Kaposi's  
CC sarcoma; blood platelets disorder such as thrombocytopaenia, leukaemia,  
CC Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders  
CC such as cancer; and protein kinase disorders such as autoimmune  
CC disorders, diabetes mellitus, psoriasis, inflammatory bowel disease,  
CC rheumatoid arthritis, and multiple sclerosis (many examples of diseases  
CC and disorders are included in the specification). The kinases, their  
CC encoding nucleic acids and antibodies are useful in screening assays,  
CC detection assays (e.g. forensic biology), and predictive medicine (e.g.  
CC diagnostic assays, prognostic assays, and monitoring clinical trials and  
CC pharmacogenomics). The kinases and their encoding nucleic acids are  
CC useful as query sequences to perform a search against public databases to  
CC identify other family members or related sequences. The present sequence  
CC encodes the kinase 12599  
XX  
SQ Sequence 24120 BP; 4719 A; 7265 C; 8278 G; 3858 T; 0 U; 0 Other;  
Query Match 100.0%; Score 385; DB 7; Length 24120;  
Best Local Similarity 100.0%; Pred. No. 9.2e-87;  
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCGAGCGAGATCTTGACATCTACGTGGTCCACCGTACTACTGCCCCCTAGGGGCTGA 60  
Db 16862 CCGAGCGAGATCTTGACATCTACGTGGTCCACCGTACTACTGCCCCCTAGGGGCTGA 16921  
QY 61 GCAGGATGCCATCACGCTCGGGAAGGCCAGTATGTGGAGGTCTCTGGATGAGCCACCC 120  
Db 16922 GCAGGATGCCATCACGCTCGGGAAGGCCAGTATGTGGAGGTCTCTGGATGAGCCACCC 16981  
QY 121 ACTGCGCTGGCTTGTCCGACCAAGCCCAAGTCCAGCCCTCACGCGAGGGTGGGT 180  
Db 16982 ACTGCGCTGGCTTGTCCGACCAAGCCCAAGTCCAGCCCTCACGCGAGGGTGGGT 17041  
QY 181 GTCACAGCTTACCTGAGCAGAGGCTCAAGTGTCACTGAGTGGGGGGCGCTGAGGC 240  
Db 17042 GTCACAGCTTACCTGAGCAGAGGCTCAAGTGTCACTGAGTGGGGGGCGCTGAGGC 17101  
QY 241 CCCTGAGTTCCTGGGAGGCTGTGTCTGAAGCAAGTATACAGGCAAGGCTGAGCTCTGT 300  
Db 17102 CCCTGAGTTCCTGGGAGGCTGTGTCTGAAGCAAGTATACAGGCAAGGCTGAGCTCTGT 17161  
QY 301 GATCCAGAGCTGTGAGTCTTGAGCAGGCCCTTCGTGGAGGAGCTGCAGTTCCTGAGAG 360  
Db 17162 GATCCAGAGCTGTGAGTCTTGAGCAGGCCCTTCGTGGAGGAGCTGCAGTTCCTGAGAG 17221  
QY 361 CCACCACTGAGCAGCCTGGAGCGC 385  
Db 17222 CCACCACTGAGCAGCCTGGAGCGC 17246

RESULT 4  
AAC74767  
ID AAC74767 standard; cDNA; 628 BP.  
XX  
AC AAC74767;





PT

XX

PS

XX

XX

XX

XX

XX

XX

Claim 1; Page 583-584; 2508pp; English.

Sequences ABP1028-ABP35561 represent 4534 novel human proteins designated ORF (open reading frame) 1-4534, and sequences ABN75054-ABN79587 represent cDNAs encoding them. The invention also encompasses polypeptides at least 80% identical to the ORF1-ORF4534 (collectively referred to as ORFX) proteins, polynucleotides at least 85% identical to the ORFX nucleic acid sequences, vectors and host cells comprising ORFX polynucleotides, the recombinant production of ORFX proteins, antibodies specific for ORFX proteins, methods of detecting ORFX polynucleotides and polypeptides, methods of screening for modulators of ORFX expression or activity, and methods of screening individuals for a predisposition to an ORFX-associated disorder. The ORFX proteins of the invention have a wide range of biological activities, such as cytokine, cell proliferation, cell differentiation, immune modulation, haematopoiesis regulation, tissue growth, angiogenesis, activin or inhibin activity, chemotactic/chemokinetic activity, haemostatic activity, thrombolytic activity, receptor/ligand, anti-inflammatory activity, tumour inhibition activity, and anti-infective activity, and may also be involved in the determination of bodily characteristics, fertility and behaviour. ORFX proteins, nucleic acids and antibodies may be used in the treatment of cancers, other proliferative disorders such as psoriasis and benign tumours, neurological disorders such as epilepsy and Alzheimer's disease, cardiovascular diseases, immune system disorders, disorders related to organ transplantation, disorders of tissue growth and regeneration, diseases such as diabetes mellitus, hypothyroidism, and cholesterol storage disease, and infectious diseases caused by viral, bacterial, fungal and other pathogens. ORFX nucleic acids may also be used as a source of primers and probes, in the detection of ORFX genomic sequences or transcripts, in the identification and cloning of homologous sequences, in genetic diagnosis, and in forensic biology. The ORFX nucleic acids may additionally be used to produce transgenic animals which may be useful for studying the function and/or activity of ORFX protein, and in drug screening. The ORFX proteins may also be used as immunogens to generate specific antibodies, which are useful in the diagnosis, treatment and monitoring of ORFX-associated diseases

XX Sequence 628 BP; 122 A; 189 C; 200 G; 116 T; 0 U; 1 Other;

Query Match 97.9%; Score 377; DB 6; Length 628;

Best Local Similarity 100.0%; Pred. No. 4.2e-85;

Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AGATCTTTGACATCTACCTGCTGACCGCTGACTACCTCCCTAGGGCTGAGCAGATG 68

DB 2 AGATCTTTGACATCTACCTGCTGACCGCTGACTACCTCCCTAGGGCTGAGCAGATG 61

QY 69 CCATCACGCTCGGGAAGGCCAGTATGTGGAGTCTTGATGTCAGCCACCCATGCGCT 128

DB 62 CCATCACGCTCGGGAAGGCCAGTATGTGGAGTCTTGATGTCAGCCACCCATGCGCT 121

QY 129 GGCTTGTCGCGACCAAGCCACCAAGTCCAGCCCTCAGCGAGGCTGGGTGTCACGAG 188

DB 122 GGCTTGTCGCGACCAAGCCACCAAGTCCAGCCCTCAGCGAGGCTGGGTGTCACGAG 181

QY 189 CTTACTCTGACAGGAGGCTCAAGCTGTCACTGTAGTGGGGGGCGGCTGAGGCCCTGAGT 248

DB 182 CTTACTCTGACAGGAGGCTCAAGCTGTCACTGTAGTGGGGGGCGGCTGAGGCCCTGAGT 241

QY 249 TCCCTGGGAGGCTGTGTGAGACGATACAGGCAAGCTGAGCTGTGTGATCAGG 308

DB 242 TCCCTGGGAGGCTGTGTGAGACGATACAGGCAAGCTGAGCTGTGTGATCAGG 301

QY 309 AGCTCTGAGTTCTGAGCAGGCGCTTCGTGGAGGAGCTGCAGTCTCTGACAGCCACCA 368

DB 302 AGCTCTGAGTTCTGAGCAGGCGCTTCGTGGAGGAGCTGCAGTCTCTGACAGCCACCA 361

QY 369 TGCAGCAGCTGAGGCGC 385

DB 362 TGCAGCAGCTGAGGCGC 378

RESULT 6

ABK84118

ID ABK84118 standard; cDNA; 4180 BP.

XX

AC ABK84118;

XX 14-AUG-2002 (first entry)

DE Human cDNA differentially expressed in granulocytic cells #689.

XX Human; ss; granulocytic cell; DNA chip; bacterial infection;

XX viral infection; parasitic infection; protozoal infection;

XX fungal infection; sterile inflammatory disease; psoriasis;

XX rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;

XX cardiac reperfusion injury; renal reperfusion injury; ARDS;

XX adult respiratory distress syndrome; inflammatory bowel disease;

XX Crohn's disease; ulcerative colitis; periodontal disease;

XX granulocyte activation; chronic inflammation; allergy.

XX Homo sapiens.

OS WO200228999-A2.

PN 11-APR-2002.

XX 03-OCT-2001; 2001WO-US030821.

XX 03-OCT-2000; 2000US-0237189P.

XX (GENE-) GENE LOGIC INC.

PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

XX WPI; 2002-435328/46.

XX Detecting granulocyte activation by detecting differential expression of

XX genes associated with granulocyte activation, which serves as diagnostic

XX markers that is useful for monitoring disease states and drug toxicity.

XX Claim 1; SEQ ID NO 689; 114pp; English.

XX The invention relates to detecting (M1) granulocyte (GC) activation

XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by

XX a DNA chip analysis as given in the specification, and comparing the

XX expression level to an expression level in an unactivated GC, where

XX differential expression of Gs is indicative of GCA. Also included are

XX modulating (M2) GCA by contacting GC with an agent that alters the

XX expression of at least one gene in Gs; (2) screening (M3) for an agent

XX capable of modulating GCA or an inflammation (especially chronic) in a

XX tissue, an allergic response in a subject, exposure of a subject to a

XX pathogen or sterile inflammatory disease using the gene expression

XX profile; (3) detecting (M4) an inflammation (especially chronic) in a

XX tissue, an allergic response in a subject, exposure of a subject to a

XX pathogen or sterile inflammatory disease, by detecting the level of

XX expression in a sample of the tissue of gene(s) from Gs, where the level

XX of expression of the gene is indicative of inflammation; (4) treating

XX (M5) an inflammation (especially chronic) or in a tissue, an allergic

XX response in a subject, exposure of a subject to a pathogen or sterile

XX inflammatory disease, by contacting a tissue having inflammation with an

XX agent that modulates the expression of gene(s) from Gs in the tissue. M1

XX is useful for detecting GCA, M2 is useful for modulating GCA; M3 is useful

XX for screening an agent capable of modulating GCA preferably in an

XX inflammation in a tissue; M4 is useful for detecting an inflammation

XX (especially chronic) in a tissue, an allergic response in a subject,

XX exposure of a subject to a pathogen or sterile inflammatory disease (e.g.

XX psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,

XX cardiac reperfusion injury, renal reperfusion injury, ARDS, adult

XX respiratory distress syndrome, inflammatory bowel disease, Crohn's

XX disease, ulcerative colitis, periodontal disease; also bacterial

XX infection, viral infection, parasitic infection, protozoal infection,

XX fungal infection and M5 is useful for treating one of the above

XX conditions. The present sequence represents a gene differentially

CC expressed in granulocytes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 4180 BP; 787 A; 1249 C; 1265 G; 859 T; 0 U; 0 Other;  
 Query Match 11.1%; Score 42.6; DB 6; Length 4180;  
 Best Local Similarity 48.5%; Pred. No. 0.67;  
 Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
 QY 144 AGCCCAAGTCCAGCCCTCAGCGAGGCTGGTGTCTACACAGCTTACCTGGACAGGA 203  
 DB 313 AGCCCGAGGAGCGAGCCGGTCCAGGAGGGGCTGGATGGCCCGCCACACAGTGGAGG 372  
 QY 204 GGCTCAAGCTGCACCTAGTGGGGGCGGCTGAGGCCCTGAGTTCTCTGGGAGGCTG 263  
 DB 373 AGCTGAAGCGCGCTGGAGGCGGGGAGCTGGAGGGCGGCGCGCTGCTGGCGCTGG 432  
 QY 264 TGTCTGAAGAGCAATACAAAGCAAGGCTGAGCTCTGTGATCCAGGAGCTGCTGAGTTCTG 323  
 DB 433 AGCGGAGCTGGCGGCGCGGCGGCGGCGGCGGCTGTGAGGAGGAGGAGTGTGGCGGC 492  
 QY 324 AGCAGGCTTGTGGAGGAGCTGAGTTCTTGGAGAGCCACCCTGAGCAGCCTGGAGC 383  
 DB 493 GCCAGAGCAAGTGGAGGCGCTGTACGAGCTGCTGCGGACCAAGTGTCTGGGCGTGTCTG 552  
 QY 384 G 384  
 DB 553 G 553  
 RESULT 7  
 ABS51830  
 ID ABS51830 standard; cDNA; 4282 BP.  
 XX  
 AC ABS51830;  
 XX  
 DT 05-NOV-2002 (first entry)  
 XX  
 DE Novel human thrombopoietin splice variant cDNA, NV-16.  
 XX  
 KW Human; gene; ss; chromosome identification; cancer; thrombopoietin;  
 KW thrombocytopenia; transporter protein; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002068342-A1.  
 XX  
 PD 06-JUN-2002.  
 XX  
 PF 08-FEB-2001; 2001US-00778927.  
 XX  
 PR 09-FEB-2000; 2000IL-00134453.  
 PR 29-MAR-2000; 2000IL-00135341.  
 XX  
 PA (KHOS/) KHOSRAVI R.  
 PA (BERN/) BERNSTEIN J.  
 XX  
 PI Khosravi R, Bernstein J;  
 XX WPI; 2002-582479/62.  
 DR P-PSDB; ABG70357.  
 XX  
 XX Novel nucleic acid sequences that code for variants obtained by  
 PT alternative splicing, homologs of known thrombopoietins, and homologs of  
 PT transporter proteins, for treating diseases or disorders e.g.  
 PT thrombocytopenia.  
 XX  
 PS Claim 1; Fig 16; 114pp; English.  
 CC The invention relates to novel nucleic acid (I) sequence of an  
 CC 15 amino acid sequence and the amino acid sequence of an encoded  
 CC 15 amino acid sequence.

CC (I). (I) is useful for a variety of diagnostic purposes and to detect and  
 CC quantitate expression of the sequences in patient's cells e.g. biopsied  
 CC tissues, for chromosome identification and for therapeutic purposes. The  
 CC novel variants also serve for detection i.e. their presence or level.  
 CC indicates disease, disorder, pathological or normal condition or  
 CC alternatively the ratio between the level of variants and the level of  
 CC original sequence from which they were varied or the ratio to other  
 CC variants may be indicative of a disease, disorder or pathological  
 CC condition. The variants are useful for the development of pharmaceuticals  
 CC for various pathological conditions in which cell cycle is not normal,  
 CC notably cancer. Thrombopoietin (TPO) product is useful in  
 CC thrombocytopenia, transporter protein (TH) product is useful in  
 CC conjunction with imaging substances for detection and imaging purposes.  
 CC The variant product, the TPO product or the TH product, its catalytic or  
 CC immunogenic fragments or oligopeptides are useful for screening  
 CC therapeutic compounds in a variety of drug screening techniques. ABS51815  
 CC .ABS51855 represent novel human splice variant coding sequences of the  
 CC invention  
 XX  
 SQ Sequence 4282 BP; 816 A; 1279 C; 1303 G; 883 T; 0 U; 1 Other;  
 Query Match 11.1%; Score 42.6; DB 6; Length 4282;  
 Best Local Similarity 48.5%; Pred. No. 0.67;  
 Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
 QY 144 AGCCCAAGTCCAGCCCTCAGCGAGGCTGGTGTCTACACAGCTTACCTGGACAGGA 203  
 DB 526 AGCCCGAGGAGCGAGCCGGTCCAGGAGGGGCTGGATGGCCCGCCACAGTGGAGG 585  
 QY 204 GGCTCAAGCTGCACCTAGTGGGGGCGGCTGAGGCCCTGAGTTCTCTGGGAGGCTG 263  
 DB 586 AGCTGAAGCGCGCTGGAGGCGGGGAGCTGGAGGCGGCGGCGGCTGCTGGCGCTGG 645  
 QY 264 TGTCTGAAGAGCAATACAAAGCAAGGCTGAGCTCTGTGATCCAGGAGCTGCTGAGTTCTG 323  
 DB 646 AGCGGAGCTGGCGGCGGCGGCGGCGGCGGCTGTGAGGAGGAGGAGTGTGGTGGCGC 705  
 QY 324 AGCAGGCTTGTGGAGGAGCTGAGTTCTTGGAGAGCCACCCTGAGCAGCCTGGAGC 383  
 DB 706 GCCAGAGCAAGTGGAGGCGCTGTACAGAGCTGCTGCGGACCAAGTGTCTGGCGGTGCTGC 765  
 QY 384 G 384  
 DB 766 G 766  
 RESULT 8  
 ABX63668  
 ID ABX63668 standard; cDNA; 4366 BP.  
 XX  
 AC ABX63668;  
 XX  
 DT 26-FEB-2003 (first entry)  
 XX  
 DE Human cDNA #668 differentially expressed in activated vascular tissue.  
 XX  
 KW Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;  
 KW hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;  
 KW gene therapy; vascular disease; cancer; coronary; artery disease;  
 KW hypertension; diabetes; pre-eclampsia; restenosis;  
 KW ischaemia-reperfusion injury; stroke.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002137081-A1.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 08-JAN-2002; 2002US-00044090.  
 XX  
 PR 28-JUL-2000; 2000US-0222469P.  
 PR 08-JAN-2001; 2001US-0260483P.  
 XX

```

Query Match      11.1%; Score 42.6; DB 7; Length 4366;
Best Local Similarity 48.5%; Fred. No. 0.67;
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0

144 AGCCACCAAGTCGACGAGGCTGACGAGGCTGGGTGTACACAGGCTACTCTGGACAGGA 203
    |||||
412 AGCCCGAGGACGACGCGGTTCAGGACAGGGGTGGATGGCCCGCCCCCACAAGTGGAGG 471

204 GGCTCAAGCTGTCACTGAGTGGGGGCGCGCTGAGGCCCTCTGAGTTCCCTGGGGAGGCTG 263
    |||
472 AGCTGAAGCGCGCTGTGAGCGCGGACAGCTGAGGCGCGCGCGCTGTCTGGCGCTGG 531

264 TGTCTGAGACGGAATACAGCCAGGCTGAGCTCTGTGATCCAGGAGCTGCTGAGTTCTTG 323
    |||
532 AGCGGGAGCTGGCGCGCGCGCGCGCGCGCTGTGACGAGGAGGAGCTGGTGC CGCG 591

324 AGCAGGCCTTCGTGGAGGAGCTGACAGTTCTCTGCAGAGCCACCACTGACGACCTGGAGC 383
    |||
592 GCCAGAGCAAGGTGAGGCGCTGTACAGCTGTCTGCGGACCAAGTGTCTGGCGGTGCTGCG 651

384 G 384
    |
652 G 652

RESULT 9
AAC84312
ID AAC84312 standard; cDNA; 1803 BP.
XX

```





XX 22-JAN-2003 (first entry)  
XX DNA topoisomerase II (TOP2) 21.34 cDNA.  
XX  
XX DNA topoisomerase II 21.34; TOP2; Gene; ss; DNA recombination; cancer;  
KW malignant tumour; haemopathy; human immunodeficiency virus; HIV;  
KW immunological disease; inflammation; development disturbance.  
XX  
XX Unidentified.  
XX  
XX Key Location/Qualifiers  
PH 48. .632  
FT /\*tag= a  
FT /product= "DNA topoisomerase II 21.34"  
XX  
XX CN1345941-A.  
XX  
XX 24-APR-2002.  
XX  
XX 29-SEP-2000; 2000CN-00125577.  
XX  
XX 29-SEP-2000; 2000CN-00125577.  
XX  
XX (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.  
XX  
XX Mao Y, Xie Y;  
XX  
XX WPI; 2002-539340/58.  
XX P-PSDB; ASG70542.  
XX  
XX New polypeptide-DNA topoisomerase II (TOP2) 21.34 for treating malignant  
PT tumor, hemopathy, development disturbance, human immunodeficiency virus  
PT infection, immunological disease and various inflammations.  
XX  
XX Claim 6; Page 26-27 (Disclosure); 34pp; Chinese.  
XX  
XX The invention relates to the polypeptide DNA topoisomerase II (TOP2)  
CC 21.34, a polynucleotide encoding the polypeptide and a method for  
CC producing the polypeptide by DNA recombination technology. The  
CC polypeptide is used for curing several diseases, such as malignant  
CC tumors, haemopathy, development disturbance, human immunodeficiency  
CC virus (HIV) infection, immunological diseases and various inflammations.  
CC This sequence represents cDNA encoding DNA topoisomerase II (TOP2) 21.34  
XX  
SQ Sequence 1302 BP; 216 A; 391 C; 422 G; 273 T; 0 U; 0 Other;  
Query Match 10.2%; Score 39.4; DB 6; Length 1302;  
Best Local Similarity 51.4%; Pred. NO. 3.3;  
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
QY 18 ACATCTAGTGTTCACCCCTGACTCTGCTCCCTAGGGGCTGACGAGTCCATCAGC 77  
DB 484 ACGCTTCGTGAGCAGCGCTAAGTGCCAGCTCTTTTGCCCTCAAGGTGCTCAACATGA 543  
QY 78 TCGCGGAAGGCCAGTATGCTGAGGTCTCTGATGAGCCACCCACTGCGTGGCTTCTCC 137  
DB 544 TGCCCGAGGAGAGTGTGTTGAGCCCTGCTGAGCCACCGAGAGCAGAGAGGCC 603  
QY 138 GCACCAACCCACCAAGTTCAGCCCTCAGCGAGGGTGGGTGTACAGCCTACC 194  
DB 604 TGGAGAAGTTGCTCCCGGCTCTTGTAGGGGCTGCTGAGCTGGTGGCCCTCC 660  
RESULT 14  
AAS26907  
ID AAS26907 standard; cDNA; 1354 BP.  
XX  
XX AAS26907;  
XX  
XX 07-NOV-2001 (first entry)  
XX  
XX Human cDNA encoding a novel secreted protein, SEQ ID 99.

XX Human; immunosuppressive; antiarthritic; ss; antitumour; cytostatic;  
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;  
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;  
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
KW preservative; antiproliferative.  
XX  
XX Homo sapiens.  
XX  
XX WO200155441-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001320.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229503P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.

PR	14-SEP-2000;	2000US-0232399P.	PR	05-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0232400P.	PR	06-DEC-2000;	2000US-0251479P.
PR	14-SEP-2000;	2000US-0232401P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251868P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251869P.
PR	14-SEP-2000;	2000US-0233065P.	PR	08-DEC-2000;	2000US-0251989P.
PR	21-SEP-2000;	2000US-0234223P.	PR	11-DEC-2000;	2000US-0251990P.
PR	21-SEP-2000;	2000US-0234274P.	PR	05-JAN-2001;	2001US-0259678P.
PR	25-SEP-2000;	2000US-0234997P.	XX	(HUMA.-) HUMAN GENOME SCI INC.	
PR	25-SEP-2000;	2000US-0234998P.	XX	Rosen CA, Barash SC, Ruben SM;	
PR	26-SEP-2000;	2000US-0235484P.	XX	WPT; 2001-476222/51.	
PR	27-SEP-2000;	2000US-0235834P.	XX	P-PSDB; AAU17002.	
PR	27-SEP-2000;	2000US-0235836P.	XX	Novel polypeptides and polynucleotides useful as diagnostic reagents to	
PR	29-SEP-2000;	2000US-0236327P.	XX	diagnose diseases or disorders associated with aberrant expression or	
PR	29-SEP-2000;	2000US-0236367P.	XX	activity of polypeptides, for treating blood clotting disorder,	
PR	29-SEP-2000;	2000US-0236368P.	XX	hemophilia.	
PR	29-SEP-2000;	2000US-0236369P.	XX	Claim 1; SEQ ID NO 99; 601pp; English.	
PR	29-SEP-2000;	2000US-0236370P.	XX	The invention relates to isolated nucleic acid molecules and their	
PR	02-OCT-2000;	2000US-0236802P.	XX	encoded secreted proteins. The nucleic acids and proteins are used	
PR	02-OCT-2000;	2000US-0237037P.	XX	prevent, treat or ameliorate a medical condition in e.g. humans, mice,	
PR	02-OCT-2000;	2000US-0237038P.	XX	rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used	
PR	02-OCT-2000;	2000US-0237039P.	XX	in diagnosing a pathological condition or susceptibility to a	
PR	02-OCT-2000;	2000US-0237040P.	XX	pathological condition. Antibodies to the proteins can also be used in	
PR	13-OCT-2000;	2000US-0239355P.	XX	immunocassays e.g. radioimmunoassays or enzyme linked immunosorbant assays	
PR	13-OCT-2000;	2000US-02393937P.	XX	(ELISA). Disorders which are diagnosed or treated include autoimmune	
PR	20-OCT-2000;	2000US-0240960P.	XX	diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.	
PR	20-OCT-2000;	2000US-02411221P.	XX	neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac	
PR	20-OCT-2000;	2000US-02411785P.	XX	arrest, cerebrovascular disorders e.g. cerebral ischaemia, angioneu-	
PR	20-OCT-2000;	2000US-0241787P.	XX	ergetic nervous system disorders e.g. Alzheimer's disease, infections caused by	
PR	20-OCT-2000;	2000US-0241808P.	XX	bacteria, viruses and fungi and ocular disorders e.g. corneal infection.	
PR	20-OCT-2000;	2000US-0241809P.	XX	and many other disorders listed in the specification. The polypeptides	
PR	20-OCT-2000;	2000US-0241826P.	XX	can also be used to aid wound healing and epithelial cell proliferation,	
PR	01-NOV-2000;	2000US-0246179P.	XX	to prevent skin aging due to sunburn, to maintain organs before	
PR	08-NOV-2000;	2000US-0246474P.	XX	transplantation, for supporting cell culture of primary tissues, to	
PR	08-NOV-2000;	2000US-0246475P.	XX	regenerate tissues and in chemotaxis. The polypeptides can also be used	
PR	08-NOV-2000;	2000US-0246476P.	XX	as a food additive or preservative to increase or decrease storage	
PR	08-NOV-2000;	2000US-0246477P.	XX	capabilities, fat content, lipid, protein, carbohydrate, vitamins,	
PR	08-NOV-2000;	2000US-0246478P.	XX	minerals, cofactors and other nutritional components. The present	
PR	08-NOV-2000;	2000US-0246523P.	XX	sequence encodes a novel secreted protein of the invention. Note: The	
PR	08-NOV-2000;	2000US-0246524P.	XX	Query Match	10.2%; Score 39.4; DB 4; Length 1354;
PR	08-NOV-2000;	2000US-0246525P.	XX	Best Local Similarity	51.4%; Pred. No. 3.3;
PR	08-NOV-2000;	2000US-0246526P.	XX	Matches	91; Conservative
PR	08-NOV-2000;	2000US-0246527P.	XX	Mismatches	0; Indels
PR	08-NOV-2000;	2000US-0246528P.	XX	Gaps	0;
PR	08-NOV-2000;	2000US-0246532P.	XX	18 ACATCTACGTGGTGCACGGCTGACTACCTCGCCCTAGGGCTGAGCAGGATGCATCACGC	77
PR	08-NOV-2000;	2000US-0246610P.	XX	511 ACGCTTCGTGAGCAGCGGCTAGTGCCAGCTCTCTTTGCCCTCAAGGTGCTCACATGA	570
PR	08-NOV-2000;	2000US-0246611P.	XX	78 TCGCGGAAGCCAGTATGTGGAGGTCTCTGATGACGCCACCCACTGCCTGGCTTGTCC	137
PR	17-NOV-2000;	2000US-0249208P.	XX	571 TGCCCGAGGAGAGCTGTTGAGGCCCTGGCTGACGCCACCGAGAAGCAAGAGGCC	630
PR	17-NOV-2000;	2000US-0249210P.	XX	138 GCACCAAGCCACCAAGTCCAGCCCTCAGCGAGGGCTGGGTGTACACAGCTACC	194
PR	17-NOV-2000;	2000US-0249211P.	XX	631 TGGAGAAGTTGCTCCCGGCTCTCTTTGAGGGCTGCTGAGTGTGGGACCTCC	687
PR	17-NOV-2000;	2000US-0249212P.	XX	RESULT 15	
PR	17-NOV-2000;	2000US-0249213P.	XX	ABQ54643	
PR	17-NOV-2000;	2000US-0249214P.	XX</		







QY 361 ACAGAGCCTCAGCCAGGAGAGGCTGCTCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 420  
DB 4638 ACAGAGCCTCAGCCAGGAGAGGCTGCTCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 4697  
QY 421 CATGCTCCAGAGCTCCTCGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCCAT 480  
DB 4698 CATGCTCCAGAGCTCCTCGAGGGCCAGGGGGCTGTTCCACAGACAGACATCTGGGCCAT 4757  
QY 481 CGGTGTGACAGCTTCATCATCTGAGCGCCGAGTACCCGGTGAAGGAGGTGACG 540  
DB 4758 CGGTGTGACAGCTTCATCATCTGAGCGCCGAGTACCCGGTGAAGGAGGTGACG 4817  
QY 541 CGACTCGAGAGAGGACTGCGCAAGGGGCTGCTCCGGCTGAGCGGCTGCTACGCGGGCT 600  
DB 4818 CGACTCGAGAGAGGACTGCGCAAGGGGCTGCTCCGGCTGAGCGGCTGCTACGCGGGCT 4877  
QY 601 GTCGGGGGCGCGCTGCTCCCTGCGCAGCACTCTGTCGCGCCAGCCCTGGGGCCGCGC 660  
DB 4878 GTCGGGGGCGCGCTGCTCCCTGCGCAGCACTCTGTCGCGCCAGCCCTGGGGCCGCGC 4937  
QY 661 CTGCGCGTCCAGCTGCTGCTGAGTGCCTGCTGAGTGCCTGAGAGGAGGCGCGCTGTCGCG 720  
DB 4938 CTGCGCGTCCAGCTGCTGCTGAGTGCCTGCTGAGTGCCTGAGAGGAGGCGCGCTGTCGCG 4997  
QY 721 GCGCGCGCTGCTGCTGCTGAGTGCCTGCTGAGTGCCTGAGAGGAGGCGCGCTGTCGCG 780  
DB 4998 GCGCGCGCTGCTGCTGCTGAGTGCCTGCTGAGTGCCTGAGAGGAGGCGCGCTGTCGCG 5057  
QY 781 GAGAGCGGCTGCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840  
DB 5058 GAGAGCGGCTGCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5117  
QY 841 GCCACACCTTGGTCTCCCGCTGCGGCTGCGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900  
DB 5118 GCCACACCTTGGTCTCCCGCTGCGGCTGCGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5177  
QY 901 CGGGCGAG 930  
DB 5178 CGGGCGAG 5207

## RESULT 2

US-10-274-978-1  
; Sequence 1, Application US/10274978  
; Patent No. 6670164  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui, et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: CLO00927-CIP-DIV  
; CURRENT APPLICATION NUMBER: US/10/274,978  
; PRIOR FILING DATE: 2002-10-22  
; PRIOR APPLICATION NUMBER: 09/858,664  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 09/711,134  
; PRIOR FILING DATE: 2000-11-14  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 5207  
; TYPE: DNA  
; ORGANISM: Human  
US-10-274-978-1

Query Match 95.4%; Score 926.8; DB 4; Length 5207;  
Best Local Similarity 99.8%; Pred. No. 1.1e-184;  
Matches 928; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CAGCGGGGGCGGCTGGCGCGCAAGATCATCCCTACCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGG 60  
DB 4278 CAGCGGGGGCGGCTGGCGCGCAAGATCATCCCTACCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGG 4337

QY 61 GCTGGCGGATACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120  
DB 4338 GCTGGCGGATACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4397  
QY 121 AGCTACCTCAGCCCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 180  
DB 4398 AGCTACCTCAGCCCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4457  
QY 181 GCTCCCTGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240  
DB 4458 GCTCCCTGCTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4517  
QY 241 GCAGATGTTGAGTGCAGCCAGTACCTGCAACACAGCAGACATCTGCAACCTGCAACCTGCAACCTG 300  
DB 4518 GCAGATGTTGAGTGCAGCCAGTACCTGCAACACAGCAGACATCTGCAACCTGCAACCTGCAACCTG 4577  
QY 301 GTCGAGAACATGATATCATACCGAATAAACAACCTGCTCAAGGTGCTGAGACCTGGGGCAATGC 360  
DB 4578 GTCGAGAACATGATATCATACCGAATAAACAACCTGCTCAAGGTGCTGAGACCTGGGGCAATGC 4637  
QY 361 ACAGAGCCTCAGCCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420  
DB 4638 ACAGAGCCTCAGCCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4697  
QY 421 CATGCTCCAGAGCTCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480  
DB 4698 CATGCTCCAGAGCTCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4757  
QY 481 CGGTGTGACAGCTTCATCATCTGAGCGCCGAGTACCCGGTGAAGGAGGTGACG 540  
DB 4758 CGGTGTGACAGCTTCATCATCTGAGCGCCGAGTACCCGGTGAAGGAGGTGACG 4817  
QY 541 CGACTCGAGAGAGGACTGCGCAAGGGGCTGCTCCGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600  
DB 4818 CGACTCGAGAGAGGACTGCGCAAGGGGCTGCTCCGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4877  
QY 601 GTCGGGGGCGCGCTGCTCCCTGCGCAGCACTCTGTCGCGCCAGCCCTGGGGCCGCGC 660  
DB 4878 GTCGGGGGCGCGCTGCTCCCTGCGCAGCACTCTGTCGCGCCAGCCCTGGGGCCGCGC 4937  
QY 661 CTGCGCGTCCAGCTGCTGCTGAGTGCCTGCTGAGTGCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 720  
DB 4938 CTGCGCGTCCAGCTGCTGCTGAGTGCCTGCTGAGTGCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4997  
QY 721 GCGCGCGCTGAGACCTTCCCTACCGCGGCTGCGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780  
DB 4998 GCGCGCGCTGAGACCTTCCCTACCGCGGCTGCGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5057  
QY 781 GAGAGCGGCTGCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840  
DB 5058 GAGAGCGGCTGCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5117  
QY 841 GCCACACCTTGGTCTCCCGCTGCGGCTGCGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900  
DB 5118 GCCACACCTTGGTCTCCCGCTGCGGCTGCGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5177  
QY 901 CGGGCGAG 930  
DB 5178 CGGGCGAG 5207

## RESULT 3

US-10-274-978-3  
; Sequence 3, Application US/10274978  
; Patent No. 6670164  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui, et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: CLO00927-CIP-DIV  
; CURRENT APPLICATION NUMBER: US/10/274,978  
; CURRENT FILING DATE: 2002-10-22

; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 5207
; TYPE: DNA
; ORGANISM: Human
US-10-274-978-3

Query Match
Best Local Similarity 95.4%; Score 926.8; DB 4; Length 5207;
Matches 928; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAGCGGGGGGGCTGGCGGCGCAAGATCATCCCTACACCCCAAGACAAGACAGCAGT 60
DB 4278 CAGCGGGGGGGCTGGCGGCGCAAGATCATCCCTACACCCCAAGACAAGACAGCAGT 4337

QY 61 GCTGGCGCAATACGAGGGCCCTCAAGGGGCTGGCGGCGCAAGACCTGGCCAGCTGCACGC 120
DB 4338 GCTGGCGCAATACGAGGGCCCTCAAGGGGCTGGCGGCGCAAGACCTGGCCAGCTGCACGC 4397

QY 121 AGCTTACCTCAGCCCGCCGACCTGGTGTCTCATCTTGAGGTGTGTCTTGGGCCGAGCT 180
DB 4398 AGCTTACCTCAGCCCGCCGACCTGGTGTCTCATCTTGAGGTGTGTCTTGGGCCGAGCT 4457

QY 181 GCTCCCTCGCTGGCGGAGAGGGCTCTCTACTCAGATCCGAGGTGAGGACTACCTGTG 240
DB 4458 GCTCCCTCGCTGGCGGAGAGGGCTCTCTACTCAGATCCGAGGTGAGGACTACCTGTG 4517

QY 241 GCAGATGTTGAGTGCCACCCAGTACTCTGCACAAACAGACACATCTCTGCACCTGGACCTGAG 300
DB 4518 GCAGATGTTGAGTGCCACCCAGTACTCTGCACAAACAGACACATCTCTGCACCTGGACCTGAG 4577

QY 301 GTCCGAGAACATGATCATCCGAATACAACTGCTCAGGTCTGGACCTGGGCAATGC 360
DB 4578 GTCCGAGAACATGATCATCCGAATACAACTGCTCAGGTCTGGACCTGGGCAATGC 4637

QY 361 ACAGAGCTTCAGCAGGAGAGGTGCTGCTCTCAGCAAGTTCAAGGACTACCTAGAGAC 420
DB 4638 ACAGAGCTTCAGCAGGAGAGGTGCTGCTCTCAGCAAGTTCAAGGACTACCTAGAGAC 4697

QY 421 CATGCTCCAGAGCTCTCTGGAGGCGCAGGGGCTTTTCCAGACAGACATCTGGGCCAT 480
DB 4698 CATGCTCCAGAGCTCTCTGGAGGCGCAGGGGCTTTTCCAGACAGACATCTGGGCCAT 4757

QY 481 CGGTGTGACAGCTTCATCATGCTGAGCGCGAGTACCCGFTGAGCAGCGAGGTGCAACG 540
DB 4758 CGGTGTGACAGCTTCATCATGCTGAGCGCGAGTACCCGFTGAGCAGCGAGGTGCAACG 4817

QY 541 CGACCTGCAGAGAGACTGCGCAAGGGGCTGTGCTGGGTGAGCCGCTCTACCGGGGCT 600
DB 4818 CGACCTGCAGAGAGACTGCGCAAGGGGCTGTGCTGGGTGAGCCGCTCTACCGGGGCT 4877

QY 601 GTCCGGGGCGCGGTGGCTTCTCTGCGAGCACTCTGTGCGCCAGCCCTGGGGCGGCG 660
DB 4878 GTCCGGGGCGCGGTGGCTTCTCTGCGAGCACTCTGTGCGCCAGCCCTGGGGCGGCG 4937

QY 661 CTGCGGCTCCAGTGTCTGCTGAGTGGCTTAAACAGAGGAGGCGCGGCTGTGTGCG 720
DB 4938 CTGCGGCTCCAGTGTCTGCTGAGTGGCTTAAACAGAGGAGGCGCGGCTGTGTGCG 4997

QY 721 GCGCGCGCGGTGACCTTCCCTACCGCGGCTGCGGCTTCTGTCGCGCAATCGCGAGAA 780
DB 4998 GCGCGCGCGGTGACCTTCCCTACCGCGGCTGCGGCTTCTGTCGCGCAATCGCGAGAA 5057

QY 781 GAGACGCGGCTGTGTATACAGAGGACAACTCTGCCCGAGGTGCGGCTGCCCGCG 840
DB 5058 GAGACGCGGCTGTGTATACAGAGGACAACTCTGCCCGAGGTGCGGCTGCCCGCG 5117

QY 841 GCCACACCCCTTGGTCTCTCCCGCTGGGGGTGCTGCTGAGACGCGGCCCAATAAAACGCCACG 900

DB 5118 GCCACACCCCTTGGTCTCTCCCGCTGGGGGTGCTGTCAGACGCGCCCAATAAAACGCACAGC 5177
QY 901 CGGGCGAGAAAAAAGAAAAAAGAAAAAAGAAAAA 930
DB 5178 CGGGCGAGAAAAAAGAAAAAAGAAAAAAGAAAAA 5207

RESULT 4
US-09-159-385-3
; Sequence 3, Application US/09159385
; Patent No. 5958748
; GENERAL INFORMATION:
; APPLICANT: AKIRA, SHIZUO
; APPLICANT: KAWAI, TARO
; TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE
; FILE REFERENCE: PH-569
; CURRENT APPLICATION NUMBER: US/09/159,385
; CURRENT FILING DATE: 1998-09-23
; EARLIER APPLICATION NUMBER: JP97/261589
; EARLIER FILING DATE: 1997-09-26
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2132
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (94)..(1455)
US-09-159-385-3

Query Match
Best Local Similarity 11.6%; Score 112.2; DB 2; Length 2132;
Matches 256; Conservative 0; Mismatches 213; Indels 6; Gaps 1;

QY 64 GCGCGAATACGAGGCGCTCAAGGGCTGGCGCACCCCGACCTGGCCAGCTGCACGAGC 123
DB 279 GCGGGAGTGAAACATCTCGGGAGATCCGGACCCCAACATCATCACCTGCACGACAT 338

QY 124 CTACCTCAGCCCCCGCACCTGCTGTCTATCTTTGGAGCTGTGTCTGGGCCCCGAGTGTCT 183
DB 339 CTTTCGAGAACAGACGAGCGTGTGCTCTCATCTCTGGAGCTGGTCTCTGGGGGGAGTCTT 398

QY 184 CCCTGCTGTCGCGAGAGGCGCTCTACTCAGATCCGAGGTGAAGGACTACCTGTGGCA 243
DB 399 TGACTTCTGCGGAGAGAGAGTGTGCTGACGAGGACGAGGCCACCCAGTTCTCAAGCA 458

QY 244 GATGTTGAGTGCCACCCAGTACTCTGCACAAACAGACATCTCTGCACCTGGACCTGAGTCT 303
DB 459 GATCTTGAGCGGCTTCACTTACCTGCACTCTAAGCGCATCGCACACTTTGACCTGAAGCC 518

QY 304 CGAGAACATGATCATCAACGAAATACACCTGTCTAA-----GGTGTGACCTGGGCAA 357
DB 519 GGAACATCATCTGCTGTCGACAAAGACGTCGCCAACCCCAACGAAATCAAGCTCATGACTT 578

QY 358 TGCACAGAGCTCAGCCAGGAGAGGTGTGCTGCTCAGACAAAGTTCAAGAGTACCTAGA 417
DB 579 CGGCATCGCCACAAGATCGAGCGGGGACGAGTTCAAGAAATCTTCGGCACCCCGGA 638

QY 418 GACCATGGTCCAGAGCTCTCTGAGGCGCCAGGGGGTGTTCACAGACATCTTGGGC 477
DB 639 GTTTGTGGCCCCAGAGATTGTGAATATGAGCGCTGGGCTTGGAGCGGACATGTGGAG 698

QY 478 CATCGGTGTGACAGCTTCTCATGCTGAGCGCGGAGTACCCGGTGAGCAGCGAG 532
DB 699 CATCGGTGTGATCACCTATATCTCTCTGAGCGGTGATCTCCCTTCTCTGGGCGAG 753

RESULT 5
US-09-186-277-3
; Sequence 3, Application US/09186277
; Patent No. 6171841

; GENERAL INFORMATION:  
 ; APPLICANT: AKIRA, SHIZUO  
 ; APPLICANT: KAWAI, TARO  
 ; TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE  
 ; FILE REFERENCE: 081356/0128  
 ; CURRENT APPLICATION NUMBER: US/09/186,277  
 ; CURRENT FILING DATE: 1998-11-05  
 ; EARLIER APPLICATION NUMBER: JP97/261589  
 ; EARLIER FILING DATE: 1997-09-26  
 ; NUMBER OF SEQ ID NOS: 8  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 3  
 ; LENGTH: 2132  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (94)..(1455)  
 US-09-186-277-3

Query Match 11.6%; Score 112.2; DB 3; Length 2132;  
 Best Local Similarity 53.9%; Pred. No. 1.2e-14;  
 Matches 256; Conservative 0; Mismatches 213; Indels 6; Gaps 1;  
 QY 64 GCGGAATACAGAGGCGCTCAAGGGGCTCGCGCACCCGACCTGGCCAGCTGACGAGC 123  
 DB 279 GCGGAGGTGAACATCTCTCGGGAGATCCGGACCCCAACATCATCACCTTGCAGCAT 338  
 QY 124 CTACCTAGCCCCCGACCTGGTGCTCATCTTGGAGCTGTGCTCTGGGCGGAGCTGCT 183  
 DB 339 CTTGAGAAACAGAGCGAGCTGGTCTCATCTCTGGAGCTGTCTCTGCGGGGAGCTCTT 398  
 QY 184 CCCCTGCTGCCGAGAGGGCTCTACTCAGATCCGAGTGAAGGACTACCTGTGGCA 243  
 DB 399 TGACTTCTCGGGGAGAAAGAGTGGCTGACGAGGAGCGAGGCCACCCAGTTCTTCAAGCA 458  
 QY 244 GATGTTAGTGCACCCAGTACCTGCACAAACAGCACATCTTGCACCTGGACCTGAGGTC 303  
 DB 459 GATCTGACGCGGTCTACTACTCTGCACTTAAGCGCATCGACACCTTGAACCTGAAGCC 518  
 QY 304 CGAGACATGATCATACCGGATACACCTGCTCAA-----GGTGTGACCTGGGCAA 357  
 DB 519 GGAATAACATCATGCTGTGGCAAGACGTCGCCCAACCCAGATCAAGCTCATCGACTT 578  
 QY 358 TGCACAGAGCCTAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGA 417  
 DB 579 CGGATCGGCACAAAGATCGAGGCGGGGAAACGAGTTCAAGAACATCTTCGGCACCCCGGA 638  
 QY 418 GACATGCTCAGAGCTCTGGAGGCGCAGGGGCTGTTCCACAGACAGACATCTGGGC 477  
 DB 639 GTTTGTGCCCCAGAGATGTGAATATGAGCCCTGGGCTTGGAGCGGACATGTGGAG 698  
 QY 478 CATCGGTGTACAGCTTTCATCTGCTGAGCGCGAGTACCCTGGTGAAGCGGAG 532  
 DB 599 CATCGGTGTACATCATATATCTCTGAGCGGTGCATCCCGCTTCTCTGGGCGAG 753

RESULT 6  
 US-09-159-385-4  
 ; Sequence 4, Application US/09159385  
 ; Patent No. 5958748  
 ; GENERAL INFORMATION:  
 ; APPLICANT: AKIRA, SHIZUO  
 ; APPLICANT: KAWAI, TARO  
 ; TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE  
 ; FILE REFERENCE: PH-569  
 ; CURRENT APPLICATION NUMBER: US/09/159,385  
 ; CURRENT FILING DATE: 1998-09-23  
 ; EARLIER APPLICATION NUMBER: JP97/261589  
 ; EARLIER FILING DATE: 1997-09-26  
 ; NUMBER OF SEQ ID NOS: 8  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 4

; LENGTH: 1429  
 ; TYPE: DNA  
 ; ORGANISM: Mus musculus  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (10)..(1353)  
 US-09-159-385-4  
 Query Match 9.7%; Score 94.2; DB 2; Length 1429;  
 Best Local Similarity 52.3%; Pred. No. 6.2e-11;  
 Matches 235; Conservative 0; Mismatches 208; Indels 6; Gaps 1;  
 QY 90 TGGGCCACCCGACCTGGCCAGCTGACGAGCCTACTCAGCCCGGACCTGGTGC 149  
 DB 221 TCGGCCACCCCAACATCATACATGTCATGAGCTTTCGAGAACAGACAGATGTGGTGC 280  
 QY 150 TCATCTTGGAGCTGTGCTCTGGGCGGAGCTGCTCCCTGCTGGCCGAGAGGCTCCT 209  
 DB 281 TGTCTCTGGAGCTGGTCTCGGTGGGAGCTTTTCGACTTCTGCGCGAGAGGAGTCA 340  
 QY 210 ACTCAGATCCGAGTGAAGACTACCTGTGGCAGATGTTGAGTGCCACCCAGTACCTGC 269  
 DB 341 TGACGAGGATGAGGCCACGACGTTCTCAAACAAATCTAGACGGTGTCCACTACCTGC 400  
 QY 270 ACAACAGCACATCTGACCTGGACCTGAGTCCGGAACATGATCATACCGAATACA 329  
 DB 401 ACTCAAGCGCATCGACACTTTGACCTGAAGCCGAGAACATCATGTGTGTCGACAAAGC 460  
 QY 330 A-----CTGTCTAAGTGTGGACCTGGGCAATGCAAGAGCCTCAGCCAGGAGAGG 383  
 DB 461 ACCGACGAGCGCCCGCATTAAGCTCATGACTTTGGCATCGCGCACAGGATCGAGGCTG 520  
 QY 384 TGCTGCCCTCAGACAGTTCAGAGACTACTAGAGCATCGCTCCAGAGCTCCTGGAGG 443  
 DB 521 GGAGCGAGTTCAAGAACATCTTTGGCACACCCGAGTTTGTGCGCCCGAGATCGTGA 580  
 QY 444 GCCAGGGGCTGTTCCACAGACAGACATCTGGGCCATCGGTGTGACAGCCTTTCATATGC 503  
 DB 581 ATGAGCCACTTGGCTTGGAGGCTGACATGTGAGCATTTGGCGTATCATCACTACATCTCC 640  
 QY 504 TGAGCGCGAGTACCCGCTGAGCAGCGAG 532  
 DB 641 TGAGCGGAGCGTCCCATTCCTTGGGCGAG 669

RESULT 7  
 US-09-186-277-4  
 ; Sequence 4, Application US/09186277  
 ; Patent No. 6171841  
 ; GENERAL INFORMATION:  
 ; APPLICANT: AKIRA, SHIZUO  
 ; APPLICANT: KAWAI, TARO  
 ; TITLE OF INVENTION: DNA CODING FOR SERINE/THREONINE KINASE  
 ; FILE REFERENCE: 081356/0128  
 ; CURRENT APPLICATION NUMBER: US/09/186,277  
 ; CURRENT FILING DATE: 1998-11-05  
 ; EARLIER APPLICATION NUMBER: JP97/261589  
 ; EARLIER FILING DATE: 1997-09-26  
 ; NUMBER OF SEQ ID NOS: 8  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1429  
 ; TYPE: DNA  
 ; ORGANISM: Mus musculus  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (10)..(1353)  
 US-09-186-277-4

Query Match 9.7%; Score 94.2; DB 3; Length 1429;  
 Best Local Similarity 52.3%; Pred. No. 6.2e-11;  
 Matches 235; Conservative 0; Mismatches 208; Indels 6; Gaps 1;

QY 90 TGGCCACCCGACCTGGCCCGGAGCTGACGAGCTACCTAGCCCGCCCGGACCTGGTGC 149  
Db 221 TCGCCACCCCAACATCATACACTGCATGACGTGTTGAGAAACAAGACAGATGGTGC 280  
QY 150 TCATCTTGGAGCTGTCTCTGGCCCGGAGCTGCTCCCTGCTGCTGGCCGAGAGGCGCTCT 209  
Db 281 TGATCCTGGAGCTGTCTCGGTGGCGAGCTTTTCGACTTCTGTGCGGAGAGGAGTCA 340  
QY 210 ACTCAGATCCGAGTGAAGACTACCTGTGGCAGATGTTAGTGCCACCCAGTACCTGC 269  
Db 341 TGACGGAGATGAGCCCGACGAGTTCCCAACAATCCTAGACGGTGTCCACTACCTGC 400  
QY 270 ACAACGACACATCTCGACCTGGACCTGAGCTCGGAGACATGATCATCAACCAATACA 329  
Db 401 ACTCAAGGATCGACACATTTGACCTGAAGCCCGAGACATCATGTTGTGGACAGC 460  
QY 330 A-----CCTGCTAAGTGTGGACCTGGGCAATGACAGACCTGACGAGAGGAGG 383  
Db 461 ACGACGACGAGCCCGCATTAAGCTCATCGACTTTGGCATCGCGACAGATCGAGGCTG 520  
QY 384 TGCTGCCCTCAGACAAGTTCAAGGACTACCTAGACCATGGCTCCAGAGCTCCTGGAGG 443  
Db 521 GCAGCGAGTTCAAGACATCTTTGGCACACCCGAGTTTGTGCGCCCGAGATCGTGA 580  
QY 444 GCCAGGGGCTTTCCACAGACAGACATCTGGGCCATCGGTGTGACAGCCTTCAATATGC 503  
Db 581 ATGAGCCACTTGGCTTGGAGGCTGACATGTGAGCAATTGGGCTCATCATCACTACATCCTCC 640  
QY 504 TGAGCGCGGAGTACCGGTGACAGCGAG 532  
Db 641 TGAGCGGAGCTCCCAATTCCTGGGCGAG 669

## RESULT 8

US-08-878-989-12

; Sequence 12, Application US/08878989

; Patent No. 5885803

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Corley, Neil C.

; APPLICANT: Guegler, Karl G.

; APPLICANT: Lal, Preeti

; APPLICANT: Goli, Surya K.

; APPLICANT: Shah, Purvi

; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN

; TITLE OF INVENTION: KINASES

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/878,989

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION NUMBER:

; APPLICATION DATA:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J J

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0321 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1282 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: PROSNOT06

; CLONE: 827431

; US-08-878-989-12

Query Match 8.6%; Score 83.2; DB 2; Length 1282;

Best Local Similarity 48.2%; Pred. No. 1.2e-08;

Matches 235; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 59 GTGCTGCGCAATACAGAGCCCTCAAGGCGCTCGCCACCCCGACCTGSCCCAGCTGCAC 118  
Db 372 GTGAGAACAGATCGCAGTGTCTCGTAGATCAGTCAACCCCAACATCGTCTCTGGAG 431  
QY 119 GCAGCTTACCTAGCCCCCGGACCTTCCACCTTACCTGGCCATGGAATGCTGAGCGGAG 178  
Db 432 GATGTCACAGAGAGCCCTTCCACCTTACCTGGCCATGGAATGCTGAGCGGAG 491  
QY 179 CTGCTCCCTGCTGCGCCGAGAGGGCTCTCTACTAGAAATCCGAGTGAAGGACTACCTG 238  
Db 492 CTGTTTACCGCATATGAGCGCGCTCTACACAGAGAAGGATGCCAGCCATCTGGTG 551  
QY 239 TGGCAGATGTTGAGTGCACCCACCTGACCAACAGCAGACATCTCGACCTGGACCTG 298  
Db 552 GGTGAGTCTTGGCGCGTCTCTACCTGACAGCCTTGGGATCGTGACCCGGACCTC 611  
QY 299 AGGTCCGAGAACATGATCATCACCAATACCACTGCTCAAGGTCTGAGACCTGGGCAAT 358  
Db 612 AAGCCCGAAACCTCTCTGATGCCACGCCCTTTTGGAGACTCGAAGATCATGGTCTCTGAC 671  
QY 359 GCACAGAGCTCAGCAGAGAGAGGTGCTGCCCTCAGACAGTTCAGGACTACCTAGAG 418  
Db 672 TTTGGACTCTTCCAAAATCCAGGCTGGGAAACATCTAGGCACCCCTGTGGACCCCTGGA 731  
QY 419 ACCATGGCTCCAGAGCTCTCGAGGGCCAGGGGCTGTTCACAGACAGACATCTGGGCC 478  
Db 732 TATGTGCCCCAGAGCTCTTGGAGCAGAAACCTACGGNAGGCCGTAGATGTGTGGCC 791  
QY 479 ATCGGTGTGACCCCTTATCATCTGAGCGCCGAGTACCCGGTGAAGCAGCGAGGTGCA 538  
Db 792 CTGGCGTCTATCTCTACATCTCTGCTGTGTGGGTACCCCTTCTACGACGAGCGAC 851  
QY 539 CGCGACCT 546  
Db 852 CTGAGCT 859

## RESULT 9

US-09-272-796-12

; Sequence 12, Application US/09272796

; Patent No. 8207148

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Corley, Neil C.

; APPLICANT: Guegler, Karl G.

; APPLICANT: Lal, Preeti

; APPLICANT: Goli, Surya K.

; APPLICANT: Shah, Purvi

; TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN

; TITLE OF INVENTION: KINASES

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/272,796  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/878,989  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0321 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1282 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSNOT06  
CLONE: 827431  
US-09-272-796-12

Query Match 8.6%; Score 83.2; DB 3; Length 1282;

Best Local Similarity 48.2%; Pred. No. 1.2e-08;

Matches 235; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY	59	GTGCTGCGCGATACGAGCCCTCAAGGGCTGCGCCACCGCAGCTGGCCAGCTGCAC	118
DB	372	GTGGAGACGAGATCGAGTGTCTCGTAGATCAGTACCCCAACATCGTCTGGAG	431
QY	119	GCAGCTACCTAGCCCCCGGACCTGCTCATCTTGGAGCTGTGTCTTGGGCCGAG	178
DB	432	GATGTCCACGAGAGCCCTCCACCTCTACCTGGCCATGGAATGTTGACGGTGGCGAG	491
QY	179	GTGCTCCCTGCTGGCGAGAGGCGCTTCTACTCAGATCCGAGTGAAGTACCTG	238
DB	492	CTGTTGACCGCATCATGAGCGCGCTCTCTACCTGACAGAGAGGATGCCAGCATCTGGTG	551
QY	239	TGGCAGATGTTGAGTGCCACCCAGTACCTGCACACACAGCATCTGCACCTGGACCTG	298
DB	552	GATCAGGTCCTTGGCGCGCTCTCTACCTGACAGAGCTGGGATCGTGCACCGGACCTC	611
QY	479	ATCGGTGACAGCCCTTATCATGTGAGCGCCGAGTACCCGGTGAAGCGAGGGGTGCA	538
DB	792	CTGGCGTCACTCTCTACATCTCTGTGTGGGTACCCCTTCTACGAGGAGCGAC	851
QY	539	CGGACCT 546	
DB	852	CCTGAGCT 859	

RESULT 10  
US-09-016-434-953  
; Sequence 953, Application US/09016434  
; Patent No. 6500938  
; GENERAL INFORMATION:  
; APPLICANT: Janice Au-Young  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
; NUMBER OF SEQUENCES: 1490  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,434  
; FILING DATE: HEREWITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0002 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 953:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1282 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: PROSNOT06  
; CLONE: 827431  
US-09-016-434-953

Query Match 8.6%; Score 83.2; DB 4; Length 1282;

Best Local Similarity 48.2%; Pred. No. 1.2e-08;

Matches 235; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY	59	GTGCTGCGCGATACGAGCCCTCAAGGGCTGCGCCACCGCAGCTGGCCAGCTGCAC	118
DB	372	GTGGAGACGAGATCGAGTGTCTCGTAGATCAGTACCCCAACATCGTCTGGAG	431
QY	119	GCAGCTACCTAGCCCCCGGACCTGCTCATCTTGGAGCTGTGTCTTGGGCCGAG	178
DB	432	GATGTCCACGAGAGCCCTCCACCTCTACCTGGCCATGGAATGTTGACGGTGGCGAG	491
QY	179	GTGCTCCCTGCTGGCGAGAGGCGCTTCTACTCAGATCCGAGTGAAGTACCTG	238
DB	492	CTGTTGACCGCATCATGAGCGCGCTCTCTACCTGACAGAGAGGATGCCAGCATCTGGTG	551
QY	239	TGGCAGATGTTGAGTGCCACCCAGTACCTGCACACACAGCATCTGCACCTGGACCTG	298
DB	552	GATCAGGTCCTTGGCGCGCTCTCTACCTGACAGAGCTGGGATCGTGCACCGGACCTC	611
QY	299	AGGTCCGAGACATGATCATCACCAGATACACCTGCTCAAGGTGCTGGAGCTGGGCAAT	358
DB	612	AAGCCCGAAACCTCTCTGTATGATGCCACCGCCCTTGGAGACTCGAAGATCATGCTCTGAC	671

QY 359 GCACAGAGCTCAGCCAGGAGAGGCTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAG 418  
Db 672 TTGGACTCTCCAAATCCAGGCTGGGAACATGCTAGGCACCGCCTGTGGGACCCCTGGA 731  
QY 419 ACATGGCTCAGAGCTCTGAGGCGCCAGGGGGCTGTTCACAGACAGACATCTGGGCC 478  
Db 732 TATGTGGCCCCAGAGCTCTTGGAGCAGAAACCTACCGGAAGGCCGTAGATGTGTGGGCC 791  
QY 479 ATCGGTGTGACAGCTTCTATCATGTGTGAGCGCCGAGTACCCGGTGTGAGCAGGAGGTGCA 538  
Db 792 CTGGGCTCATCTCTCATATCTGCTGTGTGGGTACCCCTCTTCTAGACAGAGGCAC 851  
QY 539 CCGGACCT 546  
Db 852 CCTGAGCT 859

RESULT 11  
US-09-230-896C-5  
; Sequence 5, Application US/09230896C  
; Patent No. 6635479  
; GENERAL INFORMATION:  
; APPLICANT: The Scripps Research Institute  
; APPLICANT: Sutcliffe, et al.  
; TITLE OF INVENTION: Hypothalamus-Specific Polypeptides  
; FILE REFERENCE: TSRI-548.1  
; CURRENT APPLICATION NUMBER: US/09/230,896C  
; CURRENT FILING DATE: 1999-02-02  
; PRIOR APPLICATION NUMBER: 60/023,220  
; PRIOR FILING DATE: 1996-08-02  
; NUMBER OF SEQ. ID NOS: 29  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 1458  
; TYPE: DNA  
; ORGANISM: ratus ratus  
US-09-230-896C-5

Query Match 7.9%; Score 76.6; DB 4; Length 1458;  
Best Local Similarity 50.1%; Pred. No. 2.9e-07;  
Matches 246; Conservative 0; Mismatches 239; Indels 6; Gaps 2;

QY 59 GTGCTGCCGATACGAGCCCTCAAGGGCTCGCCACCCGACCTGGCCCGCAGCTGCAC 118  
Db 221 GTGGAGATGAGATCGAGTACTCGCAGGATTAGCCACCCACATTTGGCTCTGGAG 280  
QY 119 GCAGCTTACCTAGCCCCCGGACCTGTGTCTCTTTGGAGCTGTGCTCTGGGCCCGAG 178  
Db 281 GACGTCCACGAGAGCCCTTCCCATCTCTACTTGGCCATGGAGCTGGTAACAGGTGGTAA 340  
QY 179 CTGCTCCCTGCTGCGCGAGAGGGCTCTCTACTCAGATCCGAGGTGAGGACTACCTG 238  
Db 341 CTGTTTGACGATCATGAGCGGGGCTCTACAGAGAGGATCGAGCCACCTTGTA 400  
QY 239 TGGCAGATGTTGAGTGCACCCAGTACCTGCACACACAGACATCTGCACCTGGAACCTG 298  
Db 401 GGGCAGGTCCTTGGTGTCTCTCTACCTTATAGCTGGGCTGCTGCACCGGACCTC 460  
QY 299 AGTCCGAGACATATC---ATCACCAATACACCTGTCTCAGGTGTGAGCTGGGCTGGC 355  
Db 461 AAGCTGAAACCTCTCTATGCCACACCTTTGAGGACTCCAGATCATGGTCTCTGAC 520  
QY 356 AATGCACAGAGCTCAGCCAGGAGAGGCTGTGCTGCTCAGACAAAGTTCAAGGACTACTA 415  
Db 521 TTGGCTGTCCAAATTCAGCTGGCAACATGC---TAGGACAGCTGTGGGACCCCA 577  
QY 416 GAGACCATGCTCAGAGCTCTTGGAGGCGCAGGGGGCTGTTCACAGACAGACATCTGG 475  
Db 578 GGATATGTGGCCCCAGAGCTCTCTGGAGCAGAAAACCTACCGGGAAGGCCGTAGATGTGTGG 637  
QY 476 GCATCGGTGTGACAGCTTCTATCATGTGTGAGCGCCGAGTACCCGGTGTGAGCAGGAGGT 535

Db 638 GCCTGGGTGTCATCTCTACATCTCTGTTGGGTACCCCCCTTCTATGATGAGAGC 697  
QY 536 GCACGGGACCT 546  
Db 698 GATCCTGAAT 708

RESULT 12  
US-09-016-434-1147  
; Sequence 1147, Application US/09016434  
; Patent No. 6509938  
; GENERAL INFORMATION:  
; APPLICANT: Janice Au-Young  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
; NUMBER OF SEQUENCES: 1490  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,434  
; FILING DATE: HEREWITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0002 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 1147:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2169 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GENEANK  
; CLONE: g148262  
US-09-016-434-1147

Query Match 7.3%; Score 71.2; DB 4; Length 2169;  
Best Local Similarity 49.3%; Pred. No. 4.4e-06;  
Matches 215; Conservative 0; Mismatches 218; Indels 3; Gaps 1;

QY 74 GAGGCCCTCAGGGCTGCGCCACCCGACCTGGCCCGAGCTGCACGACGCTACCTCAGC 133  
Db 253 GAGCTGCACCGAGACCTGCACGACCCGCCCATCGTSCGTTTTTCGCCACCATTTGAGGAC 312  
QY 134 CCGCGGACCTGGTGTCTCTTGGAGCTGTGCTCTGGGCGCGAGCTGCTCCCTGCCTG 193  
Db 313 GCTGACAACTACATTTCTTGGAGCTCTGCAGCCGGAAGTCCCTGGCCACATCTCG 372  
QY 194 GCGGAGGGCTCTCTACTCAGAAATCCGAGGTGAAGGACTACTGTGGCAGATGTTGAT 253  
Db 373 AAGGCCCGGACACCCCTGTGGAGCCAGAAAGTGGCTACTACTCTGGCGCAGATCTCTTCT 432  
QY 254 GCCACCCAGTACTCTGCACACGACGACATCCCTGCACCTGAGTCCGAGACATG 313

Db 433 GGCCTCAAGTACTTGCACCGGGGCTCTTGCACCGGGACCTCAAGTTGGGAAATTT 492  
Qy 314 ATCATCCGGAATCAACCTCTCAAGGTCTGGACCTGGGCAATGCACAGAGCTTCAGC 373  
Db 493 TTCACTCATGAACTGAAGTGGGATTTGGGCTGGCA---GCCGGTTG 549  
Qy 374 CAGGAGAAGTCTGCCCTCAGACAAGTTCAAGACTACCTPAGAGACCATGGTCCAGAG 433  
Db 550 GAGCTCCGGAGCAGAGGAAGAACCATCTGTGGCACCCTCAACTATGTGGCTCCAGAA 609  
Qy 434 CTCCTGGAGGCCAGGGGGCTGTTCCACAGACACATCTGGGCATCGGTGTACAGCC 493  
Db 610 GTGCTGCTGAGACAGGGCCCTGAGCGGATGTATGTCACCTGGGCTGTGTATG 669  
Qy 494 TTCACTCATGCTGAGCG 509  
Db 670 TACAGCTGCTGTGCG 685

RESULT 13  
US-08-631-097-3  
; Sequence 3, Application US/08631097  
; Patent No. 5968816  
; GENERAL INFORMATION:  
; APPLICANT: Kimchi, Adi  
; TITLE OF INVENTION: Tumor Suppressor Genes,  
; TITLE OF INVENTION: Protein Enclosed Thereby, and Use of Said Genes and Protein  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wigman, Cohen, Leitner, & Myers, P.C.  
; STREET: 900 17th Street, N.W., Suite 1000  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/631.097  
; FILING DATE: 12-Apr-96  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/11598  
; FILING DATE: 12-Oct-94  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cohen, Herbert  
; REGISTRATION NUMBER: 25,109  
; REFERENCE/DOCKET NUMBER: 0744.057  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)463-7700  
; TELEFAX: (202)473-6915  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4935 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; HYPOTHETICAL: No  
; ANTI-SENSE: No  
; FRAGMENT TYPE: No. 5968816 applicable  
; ORIGINAL SOURCE:  
; ORGANISM: homo sapiens  
; STRAIN: not applicable  
; INDIVIDUAL ISOLATE: not applicable  
; DEVELOPMENTAL STAGE: not applicable  
; TISSUE TYPE: blood  
; CELL TYPE: Leucocyte  
; CELL LINE: HeLa

; ORGANELLE: not applicable  
; IMMEDIATE SOURCE:  
; LIBRARY: not applicable  
; CLONE: not applicable  
; POSITION IN GENOME:  
; CHROMOSOME/SEGMENT: not applicable  
; MAP POSITION: not applicable  
; UNITS: not applicable  
; FEATURE:  
; NAME/KEY: Seq. ID. NO.: 3 is  
; NAME/KEY: the sequence in claim 1(iii) as Figure 8 of the specification  
; LOCATION: not available  
; IDENTIFICATION METHOD: experiment-  
; IDENTIFICATION METHOD: in specification  
; OTHER INFORMATION: prevention of IFN-2  
; OTHER INFORMATION: promoted cell death  
; PUBLICATION INFORMATION: not available  
; US-08-631-097-3  
Query Match 7.2%; Score 69.6; DB 2; Length 4935;  
Best Local Similarity 53.3%; Pred. No. 1.2e-05;  
Matches 147; Conservative 0; Mismatches 129; Indels 0; Gaps 0;  
Qy 64 GCGCGAATACGAGGGCCCTCAAGGGCTGGCGCACCCGACCTGGCCACCTGGCCAGTCGACGAGC 123  
Db 522 GGGGAGGTTCAGATCTCTGAGGAGATCCAGCACCCCAATGTCATCACCCTGCAGGAGT 581  
Qy 124 CTACTCTAGCCCCCGGACCTGGTCTCATCTTGGAGCTGTCTTGGGCCGAGCTGCT 183  
Db 582 CTATGAGAACAGACGAGCTCATCTGATCTTGAACCTCGTTCAGGTGGCAGCTGTT 641  
Qy 184 CCCCTGCTGGCGGAGAGGGCTCTCTACTCAGATCCGAGGTGAAGGACTACCTGTGGCA 243  
Db 642 TGACTTCTTAGCTGAAAGGATCTTTAACTGAAGAGGAACAATGATTTCTCAACA 701  
Qy 244 GATGTGAGTGCCACCCAGTACTGTCACAAACAGACATCTCTGACCTGGACCTGAGTGC 303  
Db 702 AATTCCTAATGGTGTACTACTGCTGCTCCCTTCAATCGCCACTTTGATCTTAAGCC 761  
Qy 304 CGAGACATGATCATCACCAGATACAACTGCTCAA 339  
Db 762 TGAGAACATAATGCTTTGGATAGAAATGTCCTCAA 797  
RESULT 14  
US-08-810-712-9  
; Sequence 9, Application US/08810712G  
; Patent No. 6160106  
; GENERAL INFORMATION:  
; APPLICANT: Yeda Research and Development Co. LTD  
; TITLE OF INVENTION: Tumor Suppressor Genes, Proteins Encoded Thereby and  
; TITLE OF INVENTION: Use of said Genes and Proteins  
; FILE REFERENCE: sequencelist  
; CURRENT APPLICATION NUMBER: US/08/810,712G  
; CURRENT FILING DATE: 1997-03-03  
; EARLIER APPLICATION NUMBER: PCT/US94/11598  
; EARLIER FILING DATE: 1994-10-12  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 5886  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (337)..(4605)  
; US-08-810-712-9  
Query Match 7.2%; Score 69.6; DB 3; Length 5886;  
Best Local Similarity 53.3%; Pred. No. 1.3e-05;  
Matches 147; Conservative 0; Mismatches 129; Indels 0; Gaps 0;  
Qy 64 GCGCGAATACGAGGGCCCTCAAGGGCTGGCGCACCCGACCTGGCCAGTCGACGAGC 123



Db 522 GCGGAGGTGAGATCTCTGAAGGAGATCCAGACCCCAATGTATCATCCCTGCACGAGGT 581  
Qy 124 CTACCTCAGCCCGCGCACCTGGTGTCTATCTTTGAGCTGTGTCTGGGCCCCGAGCTGCT 183  
Db 582 CTATGAGAACAAAGACGAGGCTATCTCTGATCTCTGGAACCTCGTTGCAGGTGGCGAGCTGTT 641  
Qy 184 CCCTGCTGCGCCAGAGGGGCTCTCTACTAGATCCGAGGTGAGGACTACCTGTGGCA 243  
Db 642 TGACTTCTTAGCTGAAAAGGAATCTTTAACTGAAGAGGAACTGAAATTTCTCAAACA 701  
Qy 244 GATGTTGAGTGCCACCCAGTACCTGCACAACCCAGCACATCTGCACCTGGACCTGAGGTC 303  
Db 702 AATTTCTTAATGGTGTCTTACTACCTGCACCTCCCTCAATCGCCACCTTTGATCTTAAGCC 761  
Qy 304 CGAGAACATGATCATACCCGAAATACACCTGCTCAA 339  
Db 762 TGAGAACATATGCTTTTGGATAGAAATGTCCCCAA 797

RESULT 15

US-09-579-664B-5  
; Sequence 5, Application US/09579664B  
; Patent No. 6514719  
; GENERAL INFORMATION:  
; APPLICANT: Immunex Corporation  
; APPLICANT: Bird, Timothy A.  
; APPLICANT: Virca, G. Duke  
; APPLICANT: Martin, Umja  
; APPLICANT: Anderson, Dirk M.  
; TITLE OF INVENTION: NOVEL MURINE AND HUMAN KINASES  
; FILE REFERENCE: 2923-A  
; CURRENT APPLICATION NUMBER: US/09/579,664B  
; CURRENT FILING DATE: 2000-05-26  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 3228  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-579-664B-5

Query Match 7.1%; Score 69; DB 4; Length 3228;  
Best Local Similarity 47.6%; Pred. No. 1.4e-05;  
Matches 204; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

Qy 60 TGCTGCGGGAATACAGGCTCTCAAGGCGCTCGCGCCACCTGGCCCGAGCTGCAAG 119  
Db 296 TCTTGACTGAGATTGAGATCTCTAAGGCAATCGGACCCCATATCGTGAGCTGAAG 355  
Qy 120 CAGCCTACCTCAGCCCCCGGCACTGGTGCTCATCTTTGGAGCTGTGTCTGGGCCCCAGC 179  
Db 356 ACTTCCAGTGGGACAAATGACAATATCTACCTCATCATGGAGTTCTGTGCAAGGGGTGACC 415  
Qy 180 TGCTCCCTGCTGCGCCGAGAGGGCTCTCTACTCAGATCCGAGGTGAAGGACTACCTGT 239  
Db 416 TGTCTCGCTTCATTCTATCCCGAGGATCTGCTGAGAGGTGGCCGCTGTTTCAAGC 475  
Qy 240 GGCAGATGTTGAGTGCCACCCAGTACCTGCACAACCCAGCACATCTCTGCACCTGGACCTGA 299  
Db 476 AGCAGTTGGCTAGTGGCTGCGAGTTCTCTGATGAGGAAACATCTCTCACTTGGATCTGA 535  
Qy 300 GGTCCGAGAACATGATCATACCGAATACACCTGCTCAAGGTGCTGGACCTGGGCAATG 359  
Db 536 AACCGCAGAACATCTCTGTGAGCTCTTTGGAGAGAGCCCACTGAAACTGGCAGACTTG 595  
Qy 360 CACAGAGCTCAGCCAGAGAGAGGTGTGCTGCCCTCAGACAAGTTCAAGGACTACCTAGAGA 419  
Db 596 GCTTTGCCAGCACATGTCCTCCGCTGGGAGGAGAAACAGTGTCTCCGTGGCTCCCGGCTCT 655  
Qy 420 CCATGGCTCCAGAGTCTCTGAGAGGCCAGGGGGCTGTTTCCACAGACACATCTGGGCCA 479  
Db 656 ATATGGCTCCTGAGATGCTGTGTCGGCGGAGATGATGCGCGTGTGGACCTCTGGTCTG 715

Qy 480 TCGGTGTGA 488  
Db 716 TCGGGGTGA 724  
Search completed: September 19, 2004, 22:12:23  
Job time : 93.2015 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

```
Run on: September 19, 2004, 19:28:00 ; Search time 502.587 Seconds
        (without alignments)
        9751.690 Million cell updates/sec
```

Title: US-10-077-130-4 COPY 23150 24120

Perfect score: 971  
Sequence: 1 cagcggcgggcgctggccg.....ggcgcataaaagtctaga 971

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0

MAXIMUM OF 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

## Listing first 45 summaries

Database : Published Applications NA:\*\*

1:	/cgn2_6/ptodata/2/pubpna/us07_PUBCOMB.seq.*
2:	/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4:	/cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq.*
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6:	/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9:	/cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10:	/cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11:	/cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
14:	/cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15:	/cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16:	/cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
17:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
18:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
19:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			ID	Description	
	Score	Match	Length			
1	971	100.0	8106	14	US-10-077-130-1	Sequence 1, Appli
2	971	100.0	24120	14	US-10-077-130-4	Sequence 4, Appli
3	946.4	97.5	4175	15	US-10-307-019-2	Sequence 2, Appli
4	926.8	95.4	5207	9	US-09-858-664A-1	Sequence 1, Appli
5	926.8	95.4	5207	13	US-10-697-263-1	Sequence 1, Appli
6	926.8	95.4	5207	13	US-10-697-263-3	Sequence 3, Appli
7	911.4	93.9	3225	17	US-10-311-034-46	Sequence 46, Appli
8	911.4	93.9	5454	13	US-10-415-011-44	Sequence 44, Appli
9	906.4	93.3	1253	13	US-10-425-114-26278	Sequence 26278, A
10	906.4	93.3	4336	13	US-10-182-243-14	Sequence 14, Appli
11	905.4	93.2	5007	15	US-10-307-019-3	Sequence 3, Appli
12	905.4	93.2	7928	15	US-10-307-019-5	Sequence 5, Appli
13	829	85.4	7893	14	US-10-077-130-3	Sequence 3, Appli
14	829	85.4	23907	14	US-10-077-130-6	Sequence 6, Appli

Query Match	Score 971;	DB 14;	Length 8106;
Best Local Similarity	100.0%;		
Best Local	100.0%;	Prod NO 1	8e-220.

Best Local Similarity 100.0%; Pred. NO: 1.8E-220;  
Matches 971: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Qv 1 CAGCGGGCGGCGTGGCCGCCAAGATCATCCCCACCCCAAGACAAGACAGCAGT 60

D<sub>b</sub> 7136 CAGCGGGGGGGGGGNTGGCGGGCAAGATCATCCCCATCCACCCCAAGGACAAGACAGCAGT 7195

## ALIGNMENTS

## RESULT 1

```

US-10-077-130-1
, Sequence 1, Application US/10077130
, Publication No. US20020168742A1
, GENERAL INFORMATION:
, APPLICANT: Kapeller-Libermann, Rosana
, APPLICANT: Acton, Susan L.
, TITLE OF INVENTION: 59079 and 12599, Protect
, TITLE OF INVENTION: Members and Uses The
, FILE REFERENCE: MPI2001-047P/RCPI(M)
, CURRENT APPLICATION NUMBER: US/10/077,130
, CURRENT FILING DATE: 2002-02-15
, PRIOR APPLICATION NUMBER: 60/369201
, PRIOR FILING DATE: 2001-02-15
, NUMBER OF SEQ ID NOS: 9
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 1
, LENGTH: 8106

```

```
QY 61 GTCGCGGAATACGAGGCGCTCAAGGGCTGCGCACCCGCACTGGCCAGCTGCAAGC 120
Db 7196 GCTGGCGGAATACGAGGCGCTCAAGGGCTGCGCACCCGCACTGGCCAGCTGCAAGC 7355
QY 121 AGCTACCTCAGCCCCCGGCACTGGTGTCTCATCTTGAGGTGTGCTGCGGCCGAGCT 180
Db 7256 AGCTACCTCAGCCCCCGGCACTGGTGTCTCATCTTGAGGTGTGCTGCGGCCGAGCT 7315
QY 181 GCTCCCTGCTGGCGAGAGGGCTCCTACTCAGAATCCAGGTGAAGGACTACCTGTG 240
Db 7316 GCTCCCTGCTGGCGAGAGGGCTCCTACTCAGAATCCAGGTGAAGGACTACCTGTG 7375
QY 241 CGAGATGTTGAGTGCCACCCAGTACCTGCACAACCCAGCACATCTCTGCACTGGACCTGAG 300
Db 7376 CGAGATGTTGAGTGCCACCCAGTACCTGCACAACCCAGCACATCTCTGCACTGGACCTGAG 7435
QY 301 GTCCGAGAACATGATCATACCGAATACAACTGCTCAAGTCTGTCGACCTGGGCAATGC 360
Db 7436 GTCCGAGAACATGATCATACCGAATACAACTGCTCAAGTCTGTCGACCTGGGCAATGC 7495
QY 361 ACAGAGCTCAGCCAGAGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 420
Db 7496 ACAGAGCTCAGCCAGAGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 7555
QY 421 CATGCTCCAGAGCTCTCGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCGCAT 480
Db 7556 CATGCTCCAGAGCTCTCGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCGCAT 7615
QY 481 CGGTGTGACAGCTTCACTCATGCTGAGCGCCGAGTACCCGGTGTGACGAGGGGTGCAAG 540
Db 7616 CGGTGTGACAGCTTCACTCATGCTGAGCGCCGAGTACCCGGTGTGACGAGGGGTGCAAG 7675
QY 541 CGACTGTGACAGAGACTGCGCAAGGGGCTGGTCCGGCTGAGCGCGCTGCTACGCGGGCT 600
Db 7676 CGACTGTGACAGAGACTGCGCAAGGGGCTGGTCCGGCTGAGCGCGCTGCTACGCGGGCT 7735
QY 601 GTCCGGGGCGCGTGGGCTTCTCTCGCGAGACACTGTGTGCGCCAGCCCTGGGGCGCGCC 660
Db 7736 GTCCGGGGCGCGTGGGCTTCTCTCGCGAGACACTGTGTGCGCCAGCCCTGGGGCGCGCC 7795
QY 661 CTGCGCTGCCAGCTGCTGCAAGTCCCGTGTGCTAACAGAGAGGGCCCGGCTGTTCGCG 720
Db 7796 CTGCGCTGCCAGCTGCTGCAAGTCCCGTGTGCTAACAGAGAGGGCCCGGCTGTTCGCG 7855
QY 721 GCCCGCGCGCTGACCTTCCCTACCGCGCGCTGCGGCTCTTCTGTCGCAATGCGAGAA 780
Db 7856 GCCCGCGCGCTGACCTTCCCTACCGCGCGCTGCGGCTCTTCTGTCGCAATGCGAGAA 7915
QY 781 GAGAGCGCGCTGTGTACAGAGAGGCAACCTGGCCCGAGTGTGCTGAGGGTGCGCCCG 840
Db 7916 GAGAGCGCGCTGTGTACAGAGAGGCAACCTGGCCCGAGTGTGCTGAGGGTGCGCCCG 7975
QY 841 GCCACACCTTGTGCTCCCGCTGGGGTGTGCTGACAGCGGCGCCCAATAAAGCCGAGC 900
Db 7976 GCCACACCTTGTGCTCCCGCTGGGGTGTGCTGACAGCGGCGCCCAATAAAGCCGAGC 8035
QY 901 CGGCGGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 960
Db 8036 CGGCGGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 8095
QY 961 AAAAGTCTAGA 971
Db 8096 AAAAGTCTAGA 8106
```

## RESULT 2

```
US-10-077-130-4
; Sequence 4, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 53079 and 12599, Protein Kinase Family
```

```
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047PIRCPI(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 24120
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(71)
; NAME/KEY: CDS
; LOCATION: (72)...(23978)
; NAME/KEY: 3'UTR
; LOCATION: (23979)...(24120)
; US-10-077-130-4
```

```
Query Match 100.0%; Score 971; DB 14; Length 24120;
Best Local Similarity 100.0%; Pred. No. 2.3e-220;
Matches 971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCGGGCGGGCGCTGGCGCCCAAGATCATCCCTTACCACCCCAAGGACAAGACAGCAGT 60
Db 23150 CAGCGGGCGGGCGCTGGCGCCCAAGATCATCCCTTACCACCCCAAGGACAAGACAGCAGT 23209
QY 61 GCTGCGGAATACGAGGGCCCTCAAGGGCTGCGCCACCGCACCTGGCCAGCTGCAAGC 120
Db 23210 GCTGCGGAATACGAGGGCCCTCAAGGGCTGCGCCACCGCACCTGGCCAGCTGCAAGC 23269
QY 121 AGCTACCTCAGCCCCCGGCACTGGTGTCTCATCTTGAGGTGTGCTTGGGCCGAGCT 180
Db 23270 AGCTACCTCAGCCCCCGGCACTGGTGTCTCATCTTGAGGTGTGCTTGGGCCGAGCT 23329
QY 181 GCTCCCTGCTGGCGAGAGGGCTCCTACTCAGAATCCGAGGTGAAGGACTACCTGTG 240
Db 23330 GCTCCCTGCTGGCGAGAGGGCTCCTACTCAGAATCCGAGGTGAAGGACTACCTGTG 23389
QY 241 GCAGATGTTGAGTGCCACCCAGTACCTGCACAACCCAGACATCTCTGCACTGGACCTGAG 300
Db 23390 GCAGATGTTGAGTGCCACCCAGTACCTGCACAACCCAGACATCTCTGCACTGGACCTGAG 23449
QY 301 GTCCGAGAACATGATCATACCGAATACAACTGCTCAAGTCTGGACCTGGGCAATGC 360
Db 23450 GTCCGAGAACATGATCATACCGAATACAACTGCTCAAGTCTGGACCTGGGCAATGC 23509
QY 361 ACAGAGCTCAGCCAGGAGAAAGGTGCTGCCCTCAGACAAGTTCAAGGACTACCTAGAGAC 420
Db 23510 ACAGAGCTCAGCCAGGAGAAAGGTGCTGCCCTCAGACAAGTTCAAGGACTACCTAGAGAC 23569
QY 421 CATGCTCCAGAGCTCTCTGGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCGAT 480
Db 23570 CATGCTCCAGAGCTCTCTGGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCGAT 23629
QY 481 CGGTGTGACAGCTTTCATCATGCTGAGCGCGAGTACCCGGTGTGAGCAGCGAGGTGCAAG 540
Db 23630 CGGTGTGACAGCTTTCATCATGCTGAGCGCGAGTACCCGGTGTGAGCAGCGAGGTGCAAG 23689
QY 541 CGACTGTGACAGAGACTGTGCGCAAGGGCTGGTTCGGGTGTGAGCGGCTGTCTACGGGGGCT 600
Db 23690 CGACTGTGACAGAGACTGTGCGCAAGGGCTGGTTCGGGTGTGAGCGGCTGTCTACGGGGGCT 23749
QY 601 GTCCGGGGCGCGCTGGGCTTCTCTGCGCAGACACTCTGTGCGCCCGAGCCCTGGGCGCGGCC 660
Db 23750 GTCCGGGGCGCGCTGGGCTTCTCTGCGCAGACACTCTGTGCGCCCGAGCCCTGGGCGCGGCC 23809
QY 661 CTGCGCGTCCAGCTGCTGCAAGTCCCTGAGTCCCTGCTAACAGAGAGGGCCCGGCTGTTCGCG 720
Db 23810 CTGCGCGTCCAGCTGCTGCAAGTCCCTGAGTCCCTGCTAACAGAGAGGGCCCGGCTGTTCGCG 23869
```

QY 721 GCCCGCGCGCTGACCTTCCCTACCGCGCGGCTGGCGCTTCTTGGTGGCAATCGAGAA 780  
Db 23870 GCCCGCGCGCTGACCTTCCCTACCGCGCGGCTGGCGCTTCTTGGTGGCAATCGAGAA 23929  
QY 781 GAGACGCGCGCTGCTGTATACAGAGGCACAACTGGCCCGAGTGGCTGAGGGTGGCCCG 840  
Db 23930 GAGACGCGCGCTGCTGTATACAGAGGCACAACTGGCCCGAGTGGCTGAGGGTGGCCCG 23989  
QY 841 GCCACACCTTGGTCTCCCGCTGGGGGTGCTGACAGCGCGCAATAAATAACGCCCGAGC 900  
Db 23990 GCCACACCTTGGTCTCCCGCTGGGGGTGCTGACAGCGCGCAATAAATAACGCCCGAGC 24049  
QY 901 CGGGCGAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 960  
Db 24050 CGGGCGAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 24109  
QY 961 AAAAGTCTAGA 971  
Db 24110 AAAAGTCTAGA 24120

RESULT 3

US-10-307-019-2  
; Sequence 2, Application US/10307019  
; Publication No. US20030108533A1  
; GENERAL INFORMATION:  
; APPLICANT: Zeng, Wenlin  
; APPLICANT: Stanton, Lawrence  
; APPLICANT: SCIOS, INC.  
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION  
; FILE REFERENCE: SCIOS 021DV1  
; CURRENT APPLICATION NUMBER: US/10/307,019  
; CURRENT FILING DATE: 2002-11-26  
; PRIOR APPLICATION NUMBER: 09/548,473  
; PRIOR FILING DATE: 2000-04-13  
; PRIOR APPLICATION NUMBER: 60/129,552  
; PRIOR FILING DATE: 1999-04-16  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 4175  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(4053)  
US-10-307-019-2

Query Match 97.5%; Score 946.4; DB 15; Length 4175;  
Best Local Similarity 99.9%; Pred. No. 1.1e-214;  
Matches 947; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGCGCGGCTGGCGCGCAAGATCATCCCTACACCCCAAGGCAAGAGCAGT 60  
Db 3228 CAGCGCGCGGCTGGCGCGCAAGATCATCCCTACACCCCAAGGCAAGAGCAGT 3287  
QY 61 GTGCGCAATACGAGGCGCTCAAGGGCTGGCGCACCGACCTGGCGCCAGCTGCAGC 120  
Db 3288 GTGCGCAATACGAGGCGCTCAAGGGCTGGCGCACCGACCTGGCGCCAGCTGCAGC 3347  
QY 121 AGCTACCTCAGCCCCCGGCACTGGTGTCTATCTTGTAGCTGTGTCTTGGGCCGAGCT 180  
Db 3348 AGCTACCTCAGCCCCCGGCACTGGTGTCTATCTTGTAGCTGTGTCTTGGGCCGAGCT 3407  
QY 181 GTTCCCTCGCTGGCGAGGCGCTTACTCAGATCCGAGGTGAAGGACTACTGTG 240  
Db 3408 GTTCCCTCGCTGGCGAGGCGCTTCTTACTCAGATCCGAGGTGAAGGACTACTGTG 3467  
QY 241 GCAGATGTTGAGTGCCACCCAGTACTGTCACCAACAGCAGCATCTTGCACCTGGACCTGAG 300  
Db 3468 GCAGATGTTGAGTGCCACCCAGTACTGTCACCAACAGCAGCATCTTGCACCTGGACCTGAG 3527  
QY 301 GTCCGAGAACATGATCATCACCAGATACAACTGCTCAAGGCTGTGGACCTGGGCAATGC 360

Db 3528 GTCCGAGAACATGATCATCACCAGATACAACTGCTCAAGGCTGTGGACCTGGGCAATGC 3587  
QY 361 ACAGAGCCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTCAAGAGACTTACCTAGAGAC 420  
Db 3588 ACAGAGCCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTCAAGAGACTTACCTAGAGAC 3647  
QY 421 CATGGTCCAGAGCTCCTGAGAGGCCAGGGGGCTGTTCACACAGACAGATCTTGGGCCAT 480  
Db 3648 CATGGTCCAGAGCTCCTGAGAGGCCAGGGGGCTGTTCACACAGACAGATCTTGGGCCAT 3707  
QY 481 CGGTGTGACAGCCTTTCATCATGCTGAGCGCGAGTACCCGGTGGAGGAGGAGGCTGACG 540  
Db 3708 CGGTGTGACAGCCTTTCATCATGCTGAGCGCGAGTACCCGGTGGAGGAGGAGGCTGACG 3767  
QY 541 CGACCTCGACAGAGGACTGCGCAAGGGGCTGTTCGGCTGAGCGCTGTGTACGCGGGCT 600  
Db 3768 CGACCTCGACAGAGGACTGCGCAAGGGGCTGTTCGGCTGAGCGCTGTGTACGCGGGCT 3827  
QY 601 GTCGGGGGCGCGCTTCTGCGCAGCACTGTGGCCCGACCTGTGGCCCGCGCCGCGCC 660  
Db 3828 GTCGGGGGCGCGCTTCTGCGCAGCACTGTGGCCCGACCTGTGGCCCGCGCCGCGCC 3887  
QY 661 CTGCGGCTCCAGCTGCTGAGTGGCGCGGTGCTGTAACAGAGGAGGCGCCGCTTGTTCGCG 720  
Db 3888 CTGCGGCTCCAGCTGCTGAGTGGCGCGGTGCTGTAACAGAGGAGGCGCCGCTTGTTCGCG 3947  
QY 721 GCGCGCGCGCTGACCTTCCCTACCGCGCGGCTGGCGCTTCTTGGTGGCAATCGCAGAA 780  
Db 3948 GCGCGCGCGCTGACCTTCCCTACCGCGCGGCTGGCGCTTCTTGGTGGCAATCGCAGAA 4007  
QY 781 GAGACGCGCGCTGCTGTACAGAGGAGCACAACCTGGCCCGAGTGGCTGAGGGTGGCCCG 840  
Db 4008 GAGACGCGCGCTGCTGTACAGAGGAGCACAACCTGGCCCGAGTGGCTGAGGGTGGCCCG 4067  
QY 841 GCCACACCTTGGTCTCCCGCTGGGGGTGCTGACAGCGCGCAATAAATAACGCCCGAGC 900  
Db 4068 GCCACACCTTGGTCTCCCGCTGGGGGTGCTGACAGCGCGCAATAAATAACGCCCGAGC 4127  
QY 901 CGGGCGAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 948  
Db 4128 CGGGCGAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4175

RESULT 4

US-09-858-664A-1  
; Sequence 1, Application US/09858664A  
; Patent No. US20020072491A1  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui, et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL00927-CIP  
; CURRENT APPLICATION NUMBER: US/09/858,664A  
; CURRENT FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 09/711,134  
; PRIOR FILING DATE: 2000-11-11  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 5207  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-858-664A-1

Query Match 95.4%; Score 926.8; DB 9; Length 5207;  
Best Local Similarity 99.8%; Pred. No. 5.1e-210;  
Matches 928; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CAGCGGCGGGCGCTGGCGCGCAAGATCATCCCTACACCCCAAGGCAAGAGCAGT 60  
Db 4278 CAGCGGCGGGCGCTGGCGCGCAAGATCATCCCTACACCCCAAGGCAAGAGCAGT 4337

Qy	61	GCTGGCGAATACGAGGCCCTCAAGGGCCTCGGCCACCCGACACTGGCCAGCTGCACGC	120
Db	438	GCTGCGGAATACGAGGCCCTCAAGGGCCTCGGCCACCCGACACTGGCCAGCTGCACGC	4397
Qy	121	AGCCTACTCAGACCCCGGACACCTGGTGCTCATCTTGGAGCTGTGCTCTTGGGCCCGAGCT	180
Db	4398	AGCCTACTCAGACCCCGGACACCTGGTGCTCATCTTGGAGCTGTGCTCTTGGGCCCGAGCT	4457
Qy	181	GCTCCCTGCTGGCGAGAGGGCTCCTACTCAGAATCCGAGTGAAGACTACCTGTG	240
Db	4458	GCTCCCTGCTGGCGAGAGGGCTCCTACTCAGAATCTGAGTGAAGACTACCTGTG	4517
Qy	241	GCAGATGTTAGTGGCACCACCTGACACCAACGACACATCCTGCACCTGGACCTGAG	300
Db	4518	GCAGATGTTAGTGGCACCACCTGACACCAACGACACATCCTGCACCTGGACCTGAG	4577
Qy	301	GTCGGAACATGATCATATCCGAATACAACTGCTCAAGTCTGGAACCTGGCAATGC	360
Db	4578	GTCGGAACATGATCATATCCGAATACAACTGCTCAAGTCTGGAACCTGGCAATGC	4637
Qy	361	ACAGAGCCTCAGCCAGGAGAAGTGTGCCCTCAGACAAGTTCAAGACTACCTTAGAGAC	420
Db	4638	ACAGAGCCTCAGCCAGGAGAAGTGTGCCCTCAGACAAGTTCAAGACTACCTTAGAGAC	4697
Qy	421	CATGGCTCCAGAGTCTCTGAGGGCCAGGGGCTGTTCCACAGACACATCTGGGGCAT	480
Db	4698	CATGGCTCCAGAGTCTCTGAGGGCCAGGGGCTGTTCCACAGACACATCTGGGGCAT	4757
Qy	481	CGGTGTGACAGCCTTCATCATGCTGAGCGCGAGTACCCGCTGAGCAGCAGGCTGCACG	540
Db	4758	CGGTGTGACAGCCTTCATCATGCTGAGCGCGAGTACCCGCTGAGCAGCAGGCTGCACG	4817
Qy	541	CGACCTCGAGAGGATGCGCAAGGGCTGGTCCGCTGAGCGCTGCTACGCGGGCT	600
Db	4818	CGACCTCGAGAGGACTGCGCAAGGGCTGGTCCGCTGAGCGCTGCTACGCGGGCT	4877
Qy	601	GTCGGGGGCGCGTGGCTCTCTGCGCAGACTCTGTGCGCCAGCCCTGGGGCGCGCC	660
Db	4878	GTCGGGGGCGCGTGGCTCTCTGCGCAGACTCTGTGCGCCAGCCCTGGGGCGCGCC	4937
Qy	661	CTGCGCTCCAGTGTCTGTAGTGCCCGTGGCTAACAGAGAGGGCCCGCTGTTTCGG	720
Db	4938	CTGCGCTCCAGTGTCTGTAGTGCCCGTGGCTAACAGAGAGGGCCCGCTGTTTCGG	4997
Qy	721	GCCCGCCCGTGACCTTCCTACCGCGCGGCTCGCGTCTTGTGGCGCAATCGCAGAA	780
Db	4998	GCCCGCCCGTGACCTTCCTACCGCGCGGCTCGCGTCTTGTGGCGCAATCGCAGAA	5057
Qy	781	GAGACGCGCTGCTGTACAAGAGGCAACAACCTGGCCAGGTGCGCTGAGGGTTCGCCCG	840
Db	5058	GAGACGCGCTGCTGTACAAGAGGCAACAACCTGGCCAGGTGCGCTGAGGGTTCGCCCG	5117
Qy	841	GCCACACCTTGTCTCCCGCTGGGGTGCCTGCGACGCGCAATAAAAAAGCCACG	900
Db	5118	GCCACACCTTGTCTCCCGCTGGGGTGCCTGCGACGCGCAATAAAAAAGCCACG	5177
Qy	901	CGGGCGAGAAAAAATAAAAAAAAAAAAAA	930
Db	5178	CGGGCGAGAAAAAATAAAAAAAAAAAAAA	5207

## RESULT 5

```

US-10-697-263-1
; Sequence 1, Application US/10697263
; Publication NO. US20040063142A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL00927-CIP-DIV2
; CURRENT APPLICATION NUMBER: US/10/697 263

```

Db 5058 GAGACGCGCTGTGTATCAAGAGGCACAACTGGCCCGAGGTGGCTGAGGGTGCCTCCG 5117  
Qy 841 GCCACACCTTGGTCTCCCGCTGGGGTGTGCTGAGAGCGCGCAATAAAAAACGCCAGC 900  
Db 5118 GCCACACCTTGGTCTCCCGCTGGGGTGTGCTGAGAGCGCGCAATAAAAAACGCCAGC 5177  
Qy 901 CGGCGAGAAAAAATAAAAAAATAAAAAA 930  
Db 5178 CGGCGAGAAAAAATAAAAAAATAAAAAA 5207

RESULT 6

US-10-697-263-3  
; Sequence 3, Application US/10697263  
; Publication No. US20040063142A1  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui, et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CLO00927-CIP-DIV2  
; CURRENT APPLICATION NUMBER: US/10697,263  
; CURRENT FILING DATE: 2003-10-31  
; PRIOR APPLICATION NUMBER: 10/274,978  
; PRIOR FILING DATE: 2002-10-22  
; PRIOR APPLICATION NUMBER: 09/858,664  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 09/711,134  
; PRIOR FILING DATE: 2000-11-14  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 5207  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-697-263-3

Query Match 95.4%; Score 926.8; DB 13; Length 5207;  
Best Local Similarity 99.8%; Pred. No. 5.1e-210;  
Matches 928; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 CAGCGGGGGGGCTGGGGCCCAAGATCATCCCTTACCACCCCAAGGCAAGAGCAGCAGT 60  
Db 4278 CAGCGGGGGGGCTGGGGCCCAAGATCATCCCTTACCACCCCAAGGCAAGAGCAGCAGT 4337  
Qy 61 GCTGGCGAATACGAGGGCCCTCAAGGGCTGGCCACCGACCTGGCCAGCTGCAGC 120  
Db 4338 GCTGGCGAATACGAGGGCCCTCAAGGGCTGGCCACCGACCTGGCCAGCTGCAGC 4397  
Qy 121 AGCCTACCTCAGCCCCCGGACCTGGTGTCTATCTTGGAGCTGTCTCTGGGCCCGAGCT 180  
Db 4398 AGCCTACCTCAGCCCCCGGACCTGGTGTCTATCTTGGAGCTGTCTCTGGGCCCGAGCT 4457  
Qy 181 GCTCCCTCGCTGGCGGAGGGCCCTCTCTCTCAGATCCGAGGTGAAGGACTACCTGTG 240  
Db 4458 GCTCCCTCGCTGGCGGAGGGCCCTCTCTCTCAGATCCGAGGTGAAGGACTACCTGTG 4517  
Qy 241 GCAGATGTTGAGTGCACCCAGTACCTGCACAAACAGACATCCCTGCACCTGGACCTGAG 300  
Db 4518 GCAGATGTTGAGTGCACCCAGTACCTGCACAAACAGACATCCCTGCACCTGGACCTGAG 4577  
Qy 301 GTCGAGAAATGATCATCACCAGATACAACTGCTCAAGGTGTGGACCTGGGCAATGC 360  
Db 4578 GTCGAGAAATGATCATCACCAGATACAACTGCTCAAGGTGTGGACCTGGGCAATGC 4637  
Qy 361 ACAGAGCTCAGCAGGAGAGGTGCTGCCCTCAGACAGTTCAAGGACTACCTAGAC 420  
Db 4638 ACAGAGCTCAGCAGGAGAGGTGCTGCCCTCAGACAGTTCAAGGACTACCTAGAC 4697  
Qy 421 CATGGCTCCAGAGCTCCTGGAGGGCCAGGGGGCTGTTCACACAGACATCTGGGCCAT 480  
Db 4698 CATGGCTCCAGAGCTCCTGGAGGGCCAGGGGGCTGTTCACACAGACATCTGGGCCAT 4757

Qy 481 CGGTGTGACAGCCTTCATCATCTGAGCGCCGAGTACCCGGTGAAGCAGGAGGTGCACG 540  
Db 4758 CGGTGTGACAGCCTTCATCATCTGAGCGCCGAGTACCCGGTGAAGCAGGAGGTGCACG 4817  
Qy 541 CGACCTGCAGAGAGGACTGGCAAGGGGCTGTGTCGGCTGAGCGCGCTGTACGCGGGCT 600  
Db 4818 CGACCTGCAGAGAGGACTGGCAAGGGGCTGTGTCGGCTGAGCGCGCTGTACGCGGGCT 4877  
Qy 601 GTCGCGGGCGCGCTGGCTTCTTCTGCGCAGCACTCTGTGCGCCCAAGCCTTGGGGCGCGGC 660  
Db 4878 GTCGCGGGCGCGCTGGCTTCTTCTGCGCAGCACTCTGTGCGCCCAAGCCTTGGGGCGCGGC 4937  
Qy 661 CTCGCGGTCCAGCTGCTGAGTGGCGGTGCTTAACAGAGAGGGCGCGCTGTTCGG 720  
Db 4938 CTCGCGGTCCAGCTGCTGAGTGGCGGTGCTTAACAGAGAGGGCGCGCTGTTCGG 4997  
Qy 721 GCCCGCGCGCTGACCTTCCCTACCGCGGGCTGCGCGCTCTTCTGTCGCAATCGCAGAA 780  
Db 4998 GCCCGCGCGCTGACCTTCCCTACCGCGGGCTGCGCGCTCTTCTGTCGCAATCGCAGAA 5057  
Qy 781 GAGACGCGCTGCTGTACAAGAGGCAACCTTGGCGCCAGGTGCGTGTGAGGGTGCCTCCG 840  
Db 5058 GAGACGCGCTGCTGTACAAGAGGCAACCTTGGCGCCAGGTGCGTGTGAGGGTGCCTCCG 5117  
Qy 841 GCCACACCTTGGTCTCCCGCTGGGGTGGCTGCGAGCGCGCAATAAAAAACGCCAGC 900  
Db 5118 GCCACACCTTGGTCTCCCGCTGGGGTGGCTGCGAGCGCGCAATAAAAAACGCCAGC 5177  
Qy 901 CGGCGGAGAAAAAATAAAAAAATAAAAAA 930  
Db 5178 CGGCGGAGAAAAAATAAAAAAATAAAAAA 5207

RESULT 7

US-10-311-034-46  
; Sequence 46, Application US/10311034  
; Publication No. US20040023242A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: YUE, Henry  
; APPLICANT: LAL, Preeti  
; APPLICANT: BANDMAN, Olga  
; APPLICANT: BOROWSKY, Mark L.  
; APPLICANT: AU-YOUNG, Janice  
; APPLICANT: LU, Yan  
; APPLICANT: GANDHI, Ameena R.  
; APPLICANT: TRIBOULEY, Catherine M.  
; APPLICANT: CHAWLA, Narinder K.  
; APPLICANT: YAO, Monique G.  
; APPLICANT: LU, Dyung Aina M.  
; APPLICANT: GREENWALD, Sara R.  
; APPLICANT: RAMKUMAR, Javalaxmi  
; APPLICANT: GRIFFIN, Jennifer A.  
; APPLICANT: KEARNEY, Liam  
; APPLICANT: BURFORD, Neil  
; APPLICANT: NGUYEN, Dannel B.  
; APPLICANT: TANG, Y. Tom  
; APPLICANT: BAUGHN, Mariah R.  
; APPLICANT: HE, Ann  
; APPLICANT: THORNTON, Michael  
; APPLICANT: HAPALIA, April  
; APPLICANT: ARVIZU, Chandra S.  
; APPLICANT: GURURAJAN, Rajagopal  
; APPLICANT: LO, Terence P.  
; APPLICANT: KHAH, Farrah A.  
; APPLICANT: RSCIPON, Shirley A.  
; APPLICANT: AZIMZAI, Yalda  
; APPLICANT: POLICKY, Jennifer L.  
; APPLICANT: DING, Li  
; APPLICANT: GREYER, Megan  
; APPLICANT: ELLIOTT, Vicki S.  
; APPLICANT: THANGAVELU, Kavitha  
; APPLICANT: BATRA, Sajeev

```
; APPLICANT: ISON, Craig H.
; TITLE OF INVENTION: HUMAN KINASES
; FILE REFERENCE: PI-0125 PCT
; CURRENT APPLICATION NUMBER: US/10/311,034
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: 60/212,073; 60/213,467; 60/215,651; 60/216,605; 60/218,372;
; 60/228,056
; PRIOR FILING DATE: 2000-06-15; 2000-06-23; 2000-06-30; 2000-07-07; 2000-07-13; 2000-0
; 25
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PERL Program
; SEQ ID NO 46
; LENGTH: 3225
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040023242A1 3016969CB1
US-10-311-034-46

Query Match          93.9%; Score 911.4; DB 17; Length 3225;
Best Local Similarity 99.9%; Pred. No. 2.1e-206;
Matches 912; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGGGGGGGCTGGCGGCGCCAGATCATCCCTTACACCCCAAGGACAGACAGT 60
DB 2313 CAGCGGGGGGGCTGGCGGCGCCAGATCATCCCTTACACCCCAAGGACAGACAGT 2372
QY 61 GCTGCGCGAATACGAGGGCCCTCAAGGGCTGCGCCACCCGACCTGGCCAGCTGACGC 120
DB 2373 GCTGCGCGAATACGAGGGCCCTCAAGGGCTGCGCCACCCGACCTGGCCAGCTGACGC 2432
QY 121 AGCTTACTCAGCCCCCGGACCTGGTGCTCATCTTGGAGTGTGCTTGGGCCGAGCT 180
DB 2433 AGCTTACTCAGCCCCCGGACCTGGTGCTCATCTTGGAGTGTGCTTGGGCCGAGCT 2492
QY 181 GCTCCCTTCCCTGGCGGAGAGGGCTCCTTACTCAGAAATCCAGGTGAAGGACTACCTGTG 240
DB 2493 GCTCCCTTCCCTGGCGGAGAGGGCTCCTTACTCAGAAATCCAGGTGAAGGACTACCTGTG 2552
QY 241 CGAGATGTTAGTGCCACCCAGTACCTGCAACAACAGACACATCTCTGACCTGGACCTGAG 300
DB 2553 CGAGATGTTAGTGCCACCCAGTACCTGCAACAACAGACACATCTCTGACCTGGACCTGAG 2612
QY 301 GTCCGAGAACATGATCATACCGAATAACAACCTGCTCAAGGTGCTGGACCTGGGCAATGC 360
DB 2613 GTCCGAGAACATGATCATACCGAATAACAACCTGCTCAAGGTGCTGGACCTGGGCAATGC 2672
QY 361 ACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 420
DB 2673 ACAGAGCTCAGCCAGGAGAGGTGCTGCCCTCAGACAAAGTTCAAGGACTACCTAGAGAC 2732
QY 421 CATGGCTCCAGAGTCTCGAGGGCCAGGGGGCTGTTCCACAGACACATCTGGGCGCAT 480
DB 2733 CATGGCTCCAGAGTCTCGAGGGCCAGGGGGCTGTTCCACAGACACATCTGGGCGCAT 2792
QY 481 CGGTGTGACAGCTTTCATCATGCTGAGCGCCGAGTACCCGGTGAAGGAGGAGGAGTGCAGC 540
DB 2793 CGGTGTGACAGCTTTCATCATGCTGAGCGCCGAGTACCCGGTGAAGGAGGAGGAGTGCAGC 2852
QY 541 CGACTGACAGAGGACTGCGAGAGGGCTGGTCCGGCTGAGCGGCTGCTACCGGGGCT 600
DB 2853 CGACTGACAGAGGACTGCGAGAGGGCTGGTCCGGCTGAGCGGCTGCTACCGGGGCT 2912
QY 601 GTCCGGGGGGCGGCTGGCTTCTTCTGCGAGACATCTGTGCGCCCGAGCCCTGGGGCCGGCC 660
DB 2913 GTCCGGGGGGCGGCTGGCTTCTTCTGCGAGACATCTGTGCGCCCGAGCCCTGGGGCCGGCC 2972
QY 661 GTGGCGTCCAGCTGCTGCGAGTGGCCGTGCTGAACAGAGAGGGCCGGCTGTTGCGG 720
DB 2973 GTGGCGTCCAGCTGCTGCGAGTGGCCGTGCTGAACAGAGAGGGCCGGCTGTTGCGG 3032
QY 721 GCCCGCGCCGCTGACCTTCCCTACCGCGCGGCTCGGGTCTTCTGTCGCAATCGCGAGAA 780
```

```
DB 3033 GCCCGCGCCGCTGACCTTCCCTACCGCGGCTGGCGGCTTCTGTCGCAATCGCGAGAA 3092
QY 781 GAGACCGCGCTGCTGTACAAAGAGGACACAACTGGCCAGGTGCGCTGAGGTGCGCCCG 840
DB 3093 GAGACCGCGCTGCTGTGTACAAAGAGGACAACTGGCCAGGTGCGCTGAGGTGCGCCCG 3152
QY 841 GCCACACCCCTTGGTCTTCCCGCTGGGGGTGCTGCGAGCGCCCAATAAAAAACGCCACGC 900
DB 3153 GCCACACCCCTTGGTCTTCCCGCTGGGGGTGCTGCGAGCGCCCAATAAAAAACGCCACGC 3212
QY 901 CGGCGGAGAAAAA 913
DB 3213 CGGCGGAGAAAAA 3225

RESULT 8
US-10-415-011-44
; Sequence 44, Application US/10415011
; Publication No. US20040053394A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: XU, Yuming
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: YAO, Monique G.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: DING, Li
; APPLICANT: TANG, Y. Tom
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: NGUYEN, Dannel B.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: LU, Yan
; APPLICANT: YUE, Henry
; APPLICANT: BURSORD, Neil
; APPLICANT: BANDEMAN, Olga
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: LAL, Preeti G.
; APPLICANT: RECIPON, Shirley A.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: BOROWSKY, Mark L.
; APPLICANT: THORNTON, Michael B.
; APPLICANT: SWARNAKER, Anita
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: KEAN, Farrán A.
; APPLICANT: ISON, Craig H.
; TITLE OF INVENTION: HUMAN KINASES
; FILE REFERENCE: PI-0262 USN
; CURRENT APPLICATION NUMBER: US/10/415,011
; CURRENT FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: PCT/US01/47728
; PRIOR FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,410
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/244,068
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60/245,708
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/247,672
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/249,565
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/252,730
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: US 60/250,807
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PERL Program
; SEQ ID NO 44
; LENGTH: 5454
```



```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040053394A1 7638121CB1
US-10-415-011-44

Query Match      93.9%; Score 911.4; DB 13; Length 5454;
Best Local Similarity 99.9%; Pred. No. 2.3e-206;
Matches 912; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGGGGGGGCTGGCGCCAGAGATCATCCCTACCAACCCCAAGACAGAGAGT 60
Db 4542 CAGCGGGGGGGCTGGCGCCAGAGATCATCCCTACCAACCCCAAGACAGAGAGT 4601

QY 61 GCTGCGGAATACGAGGCGCTCAAGGGCTCGGCCACCCGACCTGGCCAGCTGCACGC 120
Db 4602 GCTGCGGAATACGAGGCGCTCAAGGGCTCGGCCACCCGACCTGGCCAGCTGCACGC 4661

QY 121 AGCTTACTCAGCCCGCCGACCTGGTCTCATCTTGGAGTGTGCTTGGGCCGAGCT 180
Db 4662 AGCTTACTCAGCCCGCCGACCTGGTCTCATCTTGGAGTGTGCTTGGGCCGAGCT 4721

QY 181 GCTCCCTCCCTGGCCGAGAGGGCTCCTACTCAGAAATCCGAGGTGAAGACTACCTGTG 240
Db 4722 GCTCCCTCCCTGGCCGAGAGGGCTCCTACTCAGAAATCCGAGGTGAAGACTACCTGTG 4781

QY 241 GCAGATGTGAGTGCCACCCAGTACTGCAAAACAGGACATCTCTGCACTGGACCTGAG 300
Db 4782 GCAGATGTGAGTGCCACCCAGTACTGCAAAACAGGACATCTCTGCACTGGACCTGAG 4841

QY 301 GTCGAGAAATGATCATCAGCAATACAACTGCTCAAGTGTGAGACCTGGCAATGC 360
Db 4842 GTCGAGAAATGATCATCAGCAATACAACTGCTCAAGTGTGAGACCTGGCAATGC 4901

QY 361 ACAGAGCTCAGCCAGGAGAGGTGCTCCCTCAGACAAAGTTCAGGACTACCTAGAGAC 420
Db 4902 ACAGAGCTCAGCCAGGAGAGGTGCTCCCTCAGACAAAGTTCAGGACTACCTAGAGAC 4961

QY 421 CATGCTCCAGAGTCTCTGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 480
Db 4962 CATGCTCCAGAGTCTCTGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 5021

QY 481 CGGTGTGACAGCTTTCATGCTGAGCGCCGAGTACCCGCTGAGCAGCGAGGTGCAAG 540
Db 5022 CGGTGTGACAGCTTTCATGCTGAGCGCCGAGTACCCGCTGAGCAGCGAGGTGCAAG 5081

QY 541 CGACTGACAGAGAGTCTGCGAAGGGCTGGTCCGGTGAAGCGCTGTACGCGGGGCT 600
Db 5082 CGACTGACAGAGAGTCTGCGAAGGGCTGGTCCGGTGAAGCGCTGTACGCGGGGCT 5141

QY 601 GTCGGGGGGCGCGTGGGCTTCTCTGCGAGCACTCTGTGGCCCGAGCCCTGGGGCCGCC 660
Db 5142 GTCGGGGGGCGCGTGGGCTTCTCTGCGAGCACTCTGTGGCCCGAGCCCTGGGGCCGCC 5201

QY 661 CTGGCGCTCAGCTGCTGCAAGTGCCTGCTAACAGAGAGGGCCCGGCTGTTCGCG 720
Db 5202 CTGGCGCTCAGCTGCTGCAAGTGCCTGCTAACAGAGAGGGCCCGGCTGTTCGCG 5261

QY 721 GCCCGCGCGTGAACCTTCCCTACCGCGCGCTGGGCTCTTCGTGGCATTCGCGAA 780
Db 5262 GCCCGCGCGTGAACCTTCCCTACCGCGCGCTGGGCTCTTCGTGGCATTCGCGAA 5321

QY 781 GAGAGCGCGTGTGTGTAAAGAGGCAAACTGGCCGAGGTGGCTGAGGGTGCAGCCG 840
Db 5322 GAGAGCGCGTGTGTGTAAAGAGGCAAACTGGCCGAGGTGGCTGAGGGTGCAGCCG 5381

QY 841 GCCACACCTTGGTCTCCCGCTGGGGTCTGCTGAGACGCGCCAAATAAAGCCCGAC 900
Db 5382 GCCACACCTTGGTCTCCCGCTGGGGTCTGCTGAGACGCGCCAAATAAAGCCCGAC 5441

QY 901 CGGGCGAGAAAA 913
|||||
```

```
Db 5442 CGGGCGAGAAAA 5454

RESULT 9
US-10-425-114-26278
; Sequence 26278, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Zhou, Jingdong
; APPLICANT: Liu, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 26278
; LENGTH: 1253
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4119-136-D3_FLI
US-10-425-114-26278

Query Match      93.3%; Score 906.4; DB 13; Length 1253;
Best Local Similarity 99.9%; Pred. No. 2.6e-205;
Matches 907; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAGCGGGGGGGCTGGCGCCAGAGATCATCCCTACCAACCCCAAGGACAGACAGT 60
Db 346 CAGCGGGGGGGCTGGCGCCAGAGATCATCCCTACCAACCCCAAGGACAGACAGT 405

QY 61 GCTGCGGAATACGAGGCGCTCAAGGGCTGGGCCACCCGACCTGGCCAGCTGCACGC 120
Db 406 GCTGCGGAATACGAGGCGCTCAAGGGCTGGGCCACCCGACCTGGCCAGCTGCACGC 465

QY 121 AGCTTACTCAGCCCGCCGACCTGGTGTCTATCTTGGAGTGTGCTTGGGCCGAGCT 180
Db 466 AGCTTACTCAGCCCGCCGACCTGGTGTCTATCTTGGAGTGTGCTTGGGCCGAGCT 525

QY 181 GCTCCCTCCCTGGCGAGAGGGCTCTTACTCAGAAATCCGAGGTGAAGACTACCTGTG 240
Db 526 GCTCCCTCCCTGGCGAGAGGGCTCTTACTCAGAAATCCGAGGTGAAGACTACCTGTG 585

QY 241 GCAGATGTGAGTGCCACCCAGTACTGCAAAACAGACATCTCTGCACTTGAAGCTGAG 300
Db 586 GCAGATGTGAGTGCCACCCAGTACTGCAAAACAGACATCTCTGCACTTGAAGCTGAG 645

QY 301 GTCGAGAAATGATCATCAGCAATACAACTGCTCAGAGTCTGAGCTGGCAATGC 360
Db 646 GTCGAGAAATGATCATCAGCAATACAACTGCTCAGAGTCTGAGCTGGCAATGC 705

QY 361 ACAGAGCTCAGCCAGGAGAGGTGCTGGCCCTCAGAAAGTTCAGGACTACCTAGAGAC 420
Db 706 ACAGAGCTCAGCCAGGAGAGGTGCTGGCCCTCAGAAAGTTCAGGACTACCTAGAGAC 765

QY 421 CATGGCTCAGAGCTCTCGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 480
Db 766 CATGGCTCAGAGCTCTCGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 825

QY 481 CGGTGTGACAGCTTTCATCATCTGAGCGCCGAGTACCCGCTGAGCAGGAGGTGCACG 540
Db 826 CGGTGTGACAGCTTTCATCATCTGAGCGCCGAGTACCCGCTGAGCAGGAGGTGCACG 885

QY 541 CGACCTGACAGAGTCTCGCAAGGGCTGTGTCGGCTGAGCGCTGTGTACGCGGGCT 600
Db 886 CGACCTGACAGAGTCTCGCAAGGGCTGTGTCGGCTGAGCGCTGTGTACGCGGGCT 945

QY 601 GTCGGGGGGCGCGTGGGCTTCTTGTGCGAGCATCTGTGCGGCCCGCCAGCCCTGGGGCGCGCC 660
```

Db 946 GTCCGGGGCCCGTGGCTTCCTCCCTGCGAGCACTCTGTGGCCAGCCCTGGGGCCGGCC 1005  
Qy 661 CTGGCGCTCCAGCTGCTGAGTCCCGTGGCTAAACAGAGAGGGCCCGCTGTTCCGG 720  
Db 1006 CTGGCGCTCCAGCTGCTGAGTCCCGTGGCTAAACAGAGAGGGCCCGCTGTTCCGG 1065  
Qy 721 GCCCGCGCCCGTGAACCTTCCTACCGCGCGCTGCGGTCTTCGTGGCAATCGCGAGAA 780  
Db 1066 GCCCGCGCCCGTGAACCTTCCTACCGCGCGCTGCGGTCTTCGTGGCAATCGCGAGAA 1125  
Qy 781 GAGAGCGCGCTGCTGTAAAGAGGCAACCTGGCCAGAGTGGCTGAGGGTGGCCCG 840  
Db 1126 GAGAGCGCGCTGCTGTAAAGAGGCAACCTGGCCAGAGTGGCTGAGGGTGGCCCG 1185  
Qy 841 GCCACACCTTTGCTCTCCCGCTGGGTGCTGCAGACGCGCAATAAAAAACGCCAGC 900  
Db 1186 GCCACACCTTTGCTCTCCCGCTGGGTGCTGCAGACGCGCAATAAAAAACGCCAGC 1245  
Qy 901 CGGGCGAG 908  
Db 1246 CGGGCGAG 1253

## RESULT 10

US-10-182-243-14  
; Sequence 14, Application US/10182243  
; Publication No. US20040048310A1  
; GENERAL INFORMATION:  
; APPLICANT: PLOWMAN, GREGORY D.  
; APPLICANT: WHYTE, DAVID  
; APPLICANT: MANNING, GERARD  
; APPLICANT: SUDARSANAM, SUCHA  
; APPLICANT: MARTINEZ, RICARDO  
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES AND PROTEIN KINASE-LIKE  
; TITLE OF INVENTION: ENZYMES  
; FILE REFERENCE: 038602/1366  
; CURRENT APPLICATION NUMBER: US/10/182,243  
; CURRENT FILING DATE: 2003-07-07  
; PRIOR APPLICATION NUMBER: PCT/US01/02337  
; PRIOR FILING DATE: 2001-01-25  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 14  
; LENGTH: 4936  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-182-243-14

Query Match 93.3%; Score 906.4; DB 13; Length 4936;  
Best Local Similarity 99.9%; Pred. No. 3.5e-205;  
Matches 907; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 CAGCGGGCGGGCTGGCGGCAAGATCATCCCTACACCCCAAGCAAGACAGCAGT 60  
Db 4029 CAGCGGGCGGGCTGGCGGCAAGATCATCCCTACACCCCAAGCAAGACAGCAGT 4088  
Qy 61 GCTGGCGGAATACGAGGCGCTCAAGGGCTGCGCCACCCGACCTGGCCAGCTGCACGC 120  
Db 4089 GCTGGCGGAATACGAGGCGCTCAAGGGCTGCGCCACCCGACCTGGCCAGCTGCACGC 4148  
Qy 121 AGCTACTCTAGCCCCCGGACCTGGTCTCATTTGAGAGTGTGCTCTGGGCCCGAGCT 180  
Db 4149 AGCTACTCTAGCCCCCGGACCTGGTCTCATTTGAGAGTGTGCTCTGGGCCCGAGCT 4208  
Qy 181 GCTCCCTGCTGCGCGAGAGGGCTCTCTACTCAGAACTCGAGGTGAAGACTACCTGTG 240  
Db 4209 GCTCCCTGCTGCGCGAGAGGGCTCTCTACTCAGAACTCGAGGTGAAGACTACCTGTG 4268  
Qy 241 GCAGATGTTAGTGCCACCCAGTACCTGCAACACAGCAATCTCTGCACTTGAACCTTGA 300  
Db 4269 GCAGATGTTAGTGCCACCCAGTACCTGCAACACAGCAATCTCTGCACTTGAACCTTGA 4328

Qy 301 GTCCGAGAACATGATCATCACCAATACAACTGCTCAAGTCTGTGACCTGTGGCAATGC 360  
Db 4329 GTCCGAGAACATGATCATCACCAATACAACTGCTCAAGTCTGTGACCTGTGGCAATGC 4388  
Qy 361 ACAGAGCTCAGCAGGAGAGGTGCTGCTCCTCAGACAAAGTTCAAGACTACCTAGAGAC 420  
Db 4389 ACAGAGCTCAGCAGGAGAGGTGCTGCTCCTCAGACAAAGTTCAAGACTACCTAGAGAC 4448  
Qy 421 CATGCTCTCAGAGCTCCTGAGGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCAAT 480  
Db 4449 CATGCTCTCAGAGCTCCTGAGGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCAAT 4508  
Qy 481 CGGTGTGACAGCTTTCATCTGTGAGCGCCAGTACCCGCTGAGCAGGAGGTGACAGC 540  
Db 4509 CGGTGTGACAGCTTTCATCTGTGAGCGCCAGTACCCGCTGAGCAGGAGGTGACAGC 4568  
Qy 541 CGACCTGACAGAGAGTTCGCAAGGGGCTGCTCCGGCTGAGCGCTGCTACGCGGGGCT 600  
Db 4569 CGACCTGACAGAGAGTTCGCAAGGGGCTGCTCCGGCTGAGCGCTGCTACGCGGGGCT 4628  
Qy 601 GTCCGGGGCGCGGTGGCTTCTGCGCAGCACTCTGTGCGCCAGCCCTGGGGCCGGCC 660  
Db 4629 GTCCGGGGCGCGGTGGCTTCTGCGCAGCACTCTGTGCGCCAGCCCTGGGGCCGGCC 4688  
Qy 661 CTGCGCGTCCAGCTGCTGCGTGGCTAACAGAGAGGGCCCGGCTGTTGCGG 720  
Db 4689 CTGCGCGTCCAGCTGCTGCGTGGCTAACAGAGAGGGCCCGGCTGTTGCGG 4748  
Qy 721 GCCCGCGCGGTGACCTTCCCTACCGCGCGCTCGGGTCTTCTGCGCAATCGCGAGAA 780  
Db 4749 GCCCGCGCGGTGACCTTCCCTACCGCGCGCTCGGGTCTTCTGCGCAATCGCGAGAA 4808  
Qy 781 GAGAGCGCGGTGCTGTAAAGAGGCAACCTGCGCCAGTGGCTGAGGGTGGCCCG 840  
Db 4809 GAGAGCGCGGTGCTGTAAAGAGGCAACCTGCGCCAGTGGCTGAGGGTGGCCCG 4868  
Qy 841 GCCACACCTTGGTCTCCCGCTGGGGTCTGCTGCGACGCGCCCAATAAAAAACGCCAGC 900  
Db 4869 GCCACACCTTGGTCTCCCGCTGGGGTCTGCTGCGACGCGCCCAATAAAAAACGCCAGC 4928  
Qy 901 CGGGCGAG 908  
Db 4929 CGGGCGAG 4936

## RESULT 11

US-10-307-019-3  
; Sequence 3, Application US/10307019  
; Publication No. US20030108533A1  
; GENERAL INFORMATION:  
; APPLICANT: Zeng, Wenlin  
; APPLICANT: Stanton, Lawrence  
; APPLICANT: SCIOS, INC.  
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION  
; FILE REFERENCE: SCIOS-021DVI  
; CURRENT APPLICATION NUMBER: US/10/307,019  
; CURRENT FILING DATE: 2002-11-26  
; PRIOR APPLICATION NUMBER: 09/548,473  
; PRIOR FILING DATE: 2000-04-13  
; PRIOR APPLICATION NUMBER: 60/129,552  
; PRIOR FILING DATE: 1999-04-16  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 5007  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (97)...(4926)  
US-10-307-019-3

Query Match

93.2%; Score 905.4; DB 15; Length 5007;



```

QY 721 GCGCGCGCCGCTGACCTTCCCTACCGCGCGCTGCGGTCTTCTGGCGCAATCGCAGAA 780
Db 7742 GCGCGCGCCGCTGACCTTCCCTACCGCGCGCTGCGGTCTTCTGGCGCAATCGCAGAA 7801
QY 781 GAGACGCGCGCTGCTGTACAGAGGCAACACTCGGCCAGGTGCGCTGAGGGTGCCTCCG 840
Db 7802 GAGACGCGCGCTGCTGTACAGAGGCAACACTCGGCCAGGTGCGCTGAGGGTGCCTCCG 7861
QY 841 GCCACACCTTGGTCTCCCGCTCGCGGTGCTGCGAGCGGCCAATAAAGAGCCCGCAGC 900
Db 7862 GCCACACCTTGGTCTCCCGCTCGCGGTGCTGCGAGCGGCCAATAAAGAGCCCGCAGC 7921
QY 901 CGGCGCA 907
Db 7922 CGGCGCA 7928

RESULT 13
US-10-077-130-3
; Sequence 3, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047PIRCP1(N)
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US/10/077,130
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 7893
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-3

Query Match 85.4%; Score 829; DB 14; Length 7893;
Best Local Similarity 100.0%; Pred. No. 9e-187;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCGCGCGCGCTGCGCGCCAGATCATCCCTACCCACCCCAAGGACAGACAGCAGT 60
Db 7065 CAGCGCGCGCGCTGCGCGCCAGATCATCCCTACCCACCCCAAGGACAGACAGCAGT 7124
QY 61 GCTCGCGGAATACAGAGCGCTTAAAGGCGCTGCGCCACCCGCACTGGCCAGCTGCACGC 120
Db 7125 GCTCGCGGAATACAGAGCGCTTAAAGGCGCTGCGCCACCCGCACTGGCCAGCTGCACGC 7184
QY 121 AGCTTACTCAGCCCCCGGACCTGGTGCATCTTGGAGCTGTCTCTGGGCGCCAGCT 180
Db 7185 AGCTTACTCAGCCCCCGGACCTGGTGCATCTTGGAGCTGTCTCTGGGCGCCAGCT 7244
QY 181 GCTCCCTGCTGCGCGAGAGGCGCTCTTACTCAGAATCCGAGGTGAAGGACTTACCTGTG 240
Db 7245 GCTCCCTGCTGCGCGAGAGGCGCTCTTACTCAGAATCCGAGGTGAAGGACTTACCTGTG 7304
QY 241 GCAGATCTTGAAGTGCACACAGTACCTTGCACACACAGCAGACATCTTGCACCTGGACCTGAG 300
Db 7305 GCAGATCTTGAAGTGCACACAGTACCTTGCACACACAGCAGACATCTTGCACCTGGACCTGAG 7364
QY 301 GTCGAGAACATGATCATCATCCGAATACAACTGTCTCAAGGTCTGTGGAACCTGGGCAATGC 360
Db 7365 GTCGAGAACATGATCATCATCCGAATACAACTGTCTCAAGGTCTGTGGAACCTGGGCAATGC 7424
QY 361 ACAGAGCTCAGCAGAGAGAGTGTGCTCCCTCAGACAAATTCAGAGACTACCTAGAGAC 420
Db 7425 ACAGAGCTCAGCAGAGAGAGTGTGCTCCCTCAGACAAATTCAGAGACTACCTAGAGAC 7484
QY 421 CATGGCTCCAGAGCTCTCTGAGGGGCCAGGGGGGTGTTCACACAGACAGACATCTGGGCCAT 480

```

```

Db 7485 CATGGCTCCAGAGCTCTCTGAGGGGCCAGGGGCTGTTCACACAGACAGACATCTGGGCCAT 7544
QY 481 CCGTGTGACAGACCTTATCATCTGAGCGCCGAGTACCCGGTGACAGCGAGGGTGCACG 540
Db 7545 CCGTGTGACAGACCTTATCATCTGAGCGCCGAGTACCCGGTGACAGCGAGGGTGCACG 7604
QY 541 CGACCTGACAGAGAGACTGCGCAAGGGCTGCTCGGGTGAAGCCGCTGTACCGCGGGCT 600
Db 7605 CGACCTGACAGAGAGACTGCGCAAGGGCTGCTCGGGTGAAGCCGCTGTACCGCGGGCT 7664
QY 601 GTCGCGGGCGCGCTGCGCTTCTGCGAGACACTCTGTGCGCCGAGCCCTGGGCGCGGCC 660
Db 7665 GTCGCGGGCGCGCTGCGCTTCTGCGAGCACTCTGTGCGCCGAGCCCTGGGCGCGGCC 7724
QY 661 CTGCGCGCTCCAGCTGCGTGCAGTGCCTGCTAACAGAGAGAGGCGCGCGCTGTTCGCG 720
Db 7725 CTGCGCGCTCCAGCTGCGTGCAGTGCCTGCTAACAGAGAGAGGCGCGCGCTGTTCGCG 7784
QY 721 GCGCGCGCGCTGACCTTCCCTACCGCGCGCTGCGGTCTTCTGCGCAATCGCAGAA 780
Db 7785 GCGCGCGCGCTGACCTTCCCTACCGCGCGCTGCGGTCTTCTGCGCAATCGCAGAA 7844
QY 781 GAGACGCGCGCTGCTGTACAGAGGACACAACTGGCCCGAGGTGCGCTGA 829
Db 7845 GAGACGCGCGCTGCTGTACAGAGGACACAACTGGCCCGAGGTGCGCTGA 7893

RESULT 14
US-10-077-130-6
; Sequence 6, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047PIRCP1(N)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-6

Query Match 85.4%; Score 829; DB 14; Length 23907;
Best Local Similarity 100.0%; Pred. No. 1.1e-186;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAGCGCGCGCGCTGCGCGCCAGATCATCCCTACCCCAAGGACAGACAGCAGT 60
Db 23079 CAGCGCGCGCGCTGCGCGCCAGATCATCCCTACCCCAAGGACAGACAGCAGT 23138
QY 61 GCTCGCGGAATACAGAGCGCTTCAAGGGCGCTGCGCCACCCGCACTGGCCCGCAGCTGCACGC 120
Db 23139 GCTCGCGGAATACAGAGCGCTTCAAGGGCGCTGCGCCACCCGCACTGGCCCGCAGCTGCACGC 23198
QY 121 AGCTTACTCAGCCCCCGGACCTGGTGCATCTTGGAGCTGTGTCTCTGGGCGCCAGCT 180
Db 23199 AGCTTACTCAGCCCCCGGACCTGGTGCATCTTGGAGCTGTGTCTCTGGGCGCCAGCT 23258
QY 181 GCTCCCTGCTGCGCGAGAGGCGCTCTTACTCAGAATCCGAGGTGAAGGACTTACCTGTG 240
Db 23259 GCTCCCTGCTGCGCGAGAGGCGCTCTTACTCAGAATCCGAGGTGAAGGACTTACCTGTG 23318
QY 241 GCAGATGTTCAGTGTCCACCCAGTACCTGCGACAAACAGCAGACATCTGCACTGGACCTGAG 300
Db 23319 GCAGATGTTCAGTGTCCACCCAGTACCTGCGACAAACAGCAGACATCTGCACTGGACCTGAG 23378

```

Db 121 CTCCTGGAGGGCGGGGCTGTTCCACAGACACATCTGGCCATCGGTGTGACAGCC 180  
QY 494 TTTCATCATCTGAGCGCCGAGTACCCGGTGTGAGCGAGGGTGCA 538  
Db 181 TTTCATCATCTGAGCGCCGAGTACCCGGTGTGAGCGAGGGTGCA 225

Search completed: September 20, 2004, 02:51:26  
Job time : 505.587 secs

QY 301 GTCGAGACATGATCATACCGAATACAACTGCTCAAGGTCTGGAAGCTGGGCAATGC 360  
Db 23379 GTCGAGAAATGATCATACCGAATACAACTGCTCAAGGTCTGGAAGCTGGGCAATGC 23438  
QY 361 ACAGAGCCTCAGCCAGGAGAGGTGCTGCTCAGACAAAGTTCAAGGACTACCTAGAGAC 420  
Db 23439 ACAGAGCCTCAGCCAGGAGAGGTGCTGCTCAGACAAAGTTCAAGGACTACCTAGAGAC 23498  
QY 421 CATGGCTCCAGAGTCTCCTGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 480  
Db 23499 CATGGCTCCAGAGTCTCCTGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGGCCAT 23558  
QY 481 CGGTGTGACAGCCTTCATCATGTGTGAGCGCCGAGTACCCGGTGTGAGCGAGGGTGCAAG 540  
Db 23559 CGGTGTGACAGCCTTCATCATGTGTGAGCGCCGAGTACCCGGTGTGAGCGAGGGTGCAAG 23618  
QY 541 CGACCTGAGAGAGACTGCGCAGAGGGCTGCTGCGCTGAGCGCTGCTACCGGGGCT 600  
Db 23619 CGACCTGAGAGAGACTGCGCAGAGGGCTGCTGCGCTGAGCGCTGCTACCGGGGCT 23678  
QY 601 GTCGGGGGGCCGCTGGGCTTCCTGCGCAGCACTCTGTGCGCCCGAGCCCTGGGGCCGGCC 660  
Db 23679 GTCGGGGGGCCGCTGGGCTTCCTGCGCAGCACTCTGTGCGCCCGAGCCCTGGGGCCGGCC 23738  
QY 661 CTGCGCGCTCAGCTGCTGCTGAGTCCCGTGGCTAACAGAGAGGGCCCGGCTGTTGCGG 720  
Db 23739 CTGCGCGCTCAGCTGCTGCTGAGTCCCGTGGCTAACAGAGAGGGCCCGGCTGTTGCGG 23798  
QY 721 GCGCGCGCCGCTGACCTTCCCTACCGCGGCTGCGGCTCTTCTGCGCAATCGCGAGAA 780  
Db 23799 GCGCGCGCCGCTGACCTTCCCTACCGCGGCTGCGGCTCTTCTGCGCAATCGCGAGAA 23858  
QY 781 GAGAGCGCGCTGTGTGTACAGAGGCAACCTGGCCAGGTGGCTGA 829  
Db 23859 GAGAGCGCGCTGTGTGTACAGAGGCAACCTGGCCAGGTGGCTGA 23907

RESULT 15  
US-09-864-408A-2403  
; Sequence 2403, Application US/09864408A  
; Publication No. US20040009474A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Shimkets, Richard A.  
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Enc  
; FILE REFERENCE: 21402-012  
; CURRENT APPLICATION NUMBER: US/09/864,408A  
; CURRENT FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: 60/206,690  
; PRIOR FILING DATE: 2000-05-24  
; NUMBER OF SEQ ID NOS: 9068  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2403:  
; LENGTH: 225  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-864-408A-2403

Query Match 23.0%; Score 223.4; DB 11; Length 225;  
Best Local Similarity 99.6%; Pred. No. 2.1e-43;  
Matches 224; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 314 ATCATCACCAGATACAACTGCTCAAGGTCTGGGCAATGACAGAGCCTCAGC 373  
Db 1 ATCATCACCAGATACAACTGCTCAAGGTCTGGGCAATGACAGAGCCTCAGC 60  
QY 374 CAGGAGAAGGTGTGCTGCTCAGACAAAGTTCAAGGACTACCTAGAGACCATGCTCCAGAG 433  
Db 61 CAGGAGAAGGTGTGCTGCTCAGACAAAGTTCAAGGACTACCTAGAGACCATGCTCCAGAG 120  
QY 434 CTCTTGAGGGCCAGGGGCTGTTCCACAGACAGACATCTGGCCATCGGTGTGACAGCC 493

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 14:16:08 ; Search time 1977.82 Seconds  
(without alignments)  
10628.572 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_3662\_4146

Perfect score: 485  
Sequence: 1 gacgtgtacaagacggga.....ggagcagctggcacacagga 485

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.hcg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	485	100.0	20435	9	HSJ2535	AJ002535 Homo sapi
2	279.4	57.6	19475	9	HSJ14900	AJ314900 Homo sapi
3	279.4	57.6	13564	9	AL353593	AL353593 Homo sapi
4	220.6	45.5	24545	9	HSJ14898	AJ314898 Homo sapi
5	218.6	45.1	24888	6	AX883644	AX883644 Sequence
6	218.6	45.1	24888	6	BD160445	BD160445 Primer fo
7	218.6	45.1	24888	6	BD160445	BD160445 Homo sapi
8	218.6	45.1	174612	2	AC023889	AC023889 Homo sapi
9	216	44.5	10091	9	HSJ14901	AJ314901 Homo sapi
10	211.2	43.5	3956	9	AK128447	AK128447 Homo sapi
11	209.6	43.2	6400	9	HSJ14903	AJ314903 Homo sapi
12	207.4	42.8	211829	10	AL645854	AL645854 Mouse DNA
13	207.4	42.8	260998	2	AC099089	AC099089 Rattus no
14	201.2	41.5	164766	2	AC026657	AC026657 Homo sapi
15	201	41.4	44358	2	AC142478	AC142478 Rattus no
16	201	41.4	244553	2	AC098133	AC098133 Rattus no
17	180.2	37.2	25087	2	AC096931	AC096931 Rattus no
18	136.8	28.2	526	11	BV077840	BV077840 S208F6660
19	74.4	15.3	5382	9	AB014557	AB014557 Homo sapi
20	73.4	15.1	933	9	BC061909	BC061909 Homo sapi
21	73.4	15.1	1704	9	AF035292	AF035292 Homo sapi
22	73.4	15.1	3316	9	BC007201	BC007201 Homo sapi
23	71.8	14.8	2025	9	AK025946	AK025946 Homo sapi
24	70.4	14.5	5314	10	AX122340	AX122340 Mus muscu
25	67.8	14.0	724	6	AX870351	AX870351 Sequence
26	67.8	14.0	724	6	BD150413	BD150413 Primer fo
27	67.8	14.0	2184	6	AX883309	AX883309 Sequence
28	67.8	14.0	2184	6	BD160248	BD160248 Primer fo
29	67.8	14.0	2184	9	AK023854	AK023854 Homo sapi
30	67.6	13.9	3015	10	BC056205	BC056205 Mus muscu
31	66	13.6	2417	10	BC051643	BC051643 Mus muscu
32	65.8	13.6	2534	6	AX834822	AX834822 Sequence
33	65.8	13.6	2534	9	AK097489	AK097489 Homo sapi
34	65.8	13.6	2761	9	BC043916	BC043916 Homo sapi
35	63.6	13.1	5390	9	HSJ14904	AJ314904 Homo sapi
36	63.4	13.1	7218	6	I66494	I66494 Sequence 14
37	59.8	12.3	34667	9	HSJ14905	AJ314905 Homo sapi
38	59.8	12.3	164766	2	AC026657	AC026657 Homo sapi
39	59.8	12.3	174612	2	AC023889	AC023889 Homo sapi
40	57.4	11.8	572	6	AX576249	AX576249 Sequence
41	56.4	11.6	95745	9	AL359510	AL359510 Human DNA
42	55.8	11.5	225471	2	AC112361	AC112361 Rattus no
43	55.8	11.5	266167	2	AC121633	AC121633 Rattus no
44	54.8	11.3	187999	2	AC115011	AC115011 Mus muscu
45	54.6	11.3	135033	9	AC009955	AC009955 Homo sapi

ALIGNMENTS

RESULT 1  
HSAJ2535  
LOCUS  
DEFINITION Homo sapiens mRNA for obscurin (OBSN gene).  
ACCESSION AJ002535  
VERSION AJ002535.1 GI:15026973  
KEYWORDS OBSN gene; obscurin.  
SOURCE Homo sapiens (human).  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Young, P., Ehler, E. and Gautel, M.  
TITLE Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly

20435 bp mRNA linear PRI 14-SEP-2001

JOURNAL J. Cell Biol. 154 (1), 123-136 (2001)  
 MEDLINE 21342081  
 PUBMED 11448995  
 REFERENCE 2 (bases 1 to 20435)  
 AUTHORS Gautel, M.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-OCT-1997) Gautel M.S., Structural Biology Division,  
 European Molecular Biology Laboratory, Meierhofstr. 1, Heidelberg,  
 69117, GERMANY  
 REMARK Revised by author 20-JUL-2001  
 FEATURES Location/Qualifiers  
 source  
 1..20435  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="1q42.1"  
 /cell\_type="cardiomyocyte"  
 /tissue\_type="cardiac"  
 /dev\_stage="adult"  
 1..20435  
 /gene="OBSCN"  
 72..19934  
 /gene="OBSCN"  
 /function="titin binding"  
 /codon\_start=1  
 /evidence=experimental  
 /product="obscurin"  
 /protein\_id="CAC44768.1"  
 /db\_xref="GI:15026974"  
 /db\_xref="GOA:O96AA2"  
 /db\_xref="SPRTEML:O96AA2"  
 /translation="MDQPGSGAPRLTRPKAFVSVGKDATLSCOIVGNPTPOVSE  
 KDQPVTAGARFLAQGDLYRLTDLALDGSQVCRANRAGFAFAVGLQVDVRE  
 AACQAPFLLRLQTSIRVREGEATFCRVGSPAPVSNWSDKQGRRLGEGDQVRA  
 LEAGEALRRRAARPPDGGTYEYRAENPLGASAAAALVYDDNADATSRGTSTAA  
 LHLQRRRAMAEGAPAGPSTGTCTCTTEGRIARLSVYTGPKETVMKGGQ  
 LVTEGRVHVDAQENFLKLPCKOSDRGLTCTASNLVGYTSGVLVWREPAVP  
 PKRLDLEVREKESATFLCEVPQSTEAAMFKEETRLWASAKYGTIEEGTETRLVR  
 NVSADDAVVICETPEGSRVAELAQGNLLRLPKTRVAVGDTAMFCVELAVPVGP  
 VHLRQPEVVAGVRMAISSVILSNPPTPHGERPVTIDGVLVEKKLXGLVTRICHEA  
 EWAPETLTVDAEENGFQFVSALNSFGQSPYLEPFGPHLAPLXALVRLPLKVAQ  
 VEGEVTFSDVLTVASAGFWLDQALKASVYVEIHCDRTRHTLTIREVPASLHGAQ  
 KFAVNGTSSIRMEVRAAPGLTANKPPAAARAEVRLARLHEEAOLLAELSDQAAATWL  
 KDGTLSGPKYEVQASAGRRVLLVRDVARDDAGLYECVSRGGRIAVOLSVQGLARFL  
 HKMAGSGDAVAGGPAQFCETSEAHVHVHVHVKDGMELHSGSERFLQEDVGTRELV  
 AATVTRDDEGTSCRVGDSVDRLRLRVSEPKVFAKEQLARRKLAQAGASATLSCV  
 AQTETVTKDGLVSSSKVMEATGCTRRLLVQQAGADAGEYSCAGGGRUSFFH  
 LDVKEPWFVAXQVHSEVQAEAGANATLSCEVAQAQAEVMMYKDGKLSLKVHY  
 EAKGRRLVQQAGKTADGDCYCEARGQRVSFRLHITPEPMFAKEQSVHNEVQEA  
 GASAMLSCEVAQATEVTWKDGKLSLSSKVGMEVKGCTRRLLVPAQKADAGEYSC  
 BAGQSVFHLHTPEPKGVFAKQSVHNEVQAEAGTTAMLSCVAPQOTETVWYKDGK  
 KLSSSKVRMEVKGCTRLVQQVQKADAGEYSCAGGQRVSFQLHITPEKAVFAKEQ  
 LVHNEVTEAGASATLSCEVAQAQTEVTWKDGKLSLSSSKVRIRIEAGCKRVLVQQA  
 QVADAGYTEAGQRUSFLHDSPEKPAFAKESQLARRKVAQAEAIATLSCEVAQAQ  
 TEVTWKDGKLSLSSSKVRMEVAGCTRRLLVQQACQADAGEYSCAGGQRUSFLDVA  
 EPKVFVFAKEQVHREVAQAQATLSCVQAQTEVTWKDGKLSLSSSKVRIRIEAGCK  
 CTRLLVQQAGADAGEYSCAGGQRUSFLHVAEPKAVFAKEQSPASREVQAEAGTSA  
 TLSCVQAQATEVTWKDGKLSLSSSKVRMEVAGCTRRLLVQDAGADAGEYSCAGD  
 QRLSFLHVAEPKVPFAKEQPAHREVAQAQAEAGATLSCVQAQTEVTWKDGKLSL  
 SSKVRVAVGCTRLVQQAGADAGEYSCAGGQRUSFLRLHVAELPEQISERPCHRE  
 PLVVKHEILITATLTPSAATVWLKQGEIIRRSKRHETASQGHITLTHGAQVL  
 DSALYSKRVGAEGDPPVQVEEAAKFCRLLEPVCVGGELGVTTLACELSPACAEVVR  
 CGMTQPRVGRFQMAEGPVRLTILGLAEADAGEYSCERDDHTSAQTSVPRVYK  
 FMGSLTVAEEGEATFCQVSPSVAVVWFRDGLALPQSEKFAISQSGASHLSTIS  
 DLVLEDAGQITVAEAGSSAALVREAPVLPKLEPOTVEERSVSTLEVLTRTPW  
 ELKTRNATAPAKNVEIHAEGARHLVLHNVGFADRGFGCETPDDTKTAKLTVM  
 ROVRLVRLQAVAREOGCTATMEVQLSHADVDSWTRDGLRQQQGTCHLAVGPMHT  
 LITSLGRPEDSGLVFAEGVTHSAKLIVTELPVSRPLVQVQGTTEKEKLTCELELS  
 RPNVWRLKLDYRLRAEDAGEIQAAGACRSRLTIYRCEFADQGVVCDADHQAQSSAV  
 KVQGRYTLTKVRLADAGEIQAFAENAESRQLRVELPVLVRLPDKTAMEKHR  
 GVLECCQVSASQAQVRFKSGQELQPGPKVELVSDGLYRLKLLTSDVHAEDEDTYTCDAG

DVKTSAQFFVEEQSITIVRGLQDVTVMPEPAPAFECETSIIPSVRPPKWLKGLTKVLOAG  
 GNVLEQEGTVHRLMLRRTCTMTGCPVHFTVKGSRGARLVSDIPVLTPLPPEKTCG  
 RELQSVLSCDFRPAPKAVQWYKDTPLSPSEKFKMSLEGOMAEELRLMLPADAGVY  
 RCOAGSASSTEVTVAEAREVTVTGPLODABATEEGWASFCSEGLHEDEEWSLGNMP  
 LYNDSFHEISHKRRHTLVLSIQRADAGIVRASLSLXVTSARLAEVGVKPVFLKALD  
 DLSAREGTLALQCEVDEPAHVVRKQDQVQMGDFSKYDFLHTATRGVLVHVDVSPED  
 AGLYTCHVGSBETARVRVHDLHVTIKRLTQMVEVLEGECSFCVLSHESASPAW  
 TVGKTVGSSFRQATRCGRKYIILVREAPSDAGEVVSFVGRVTSKASITVRRPAA  
 IIKPLEDQWASGFDELRCRSLRAGTVPVHMKDKRKAKIRKSGKYDVVCEGTMLAVR  
 GASLKADAGEYTCVEASASTASLHAEKANCITEELTNLOVEEKGTAVFTCKTSHPPAA  
 TVTWKGLLELRASQKQPSOEGTLRLTISALEKADSDTVTCIDIGAOQASRKLTVQ  
 RVVHILEDVDVQEGSSATFCRI SPANVEPVHFLDKTPLHANELEIDADPGGY  
 HVTLRLQALKDSGTIIFEAGDQASAAALRVETKXPSVFSRELTDATITNEBEDTLVGE  
 TSTCDIPMCVTKDGTURKSARCSQLSHGHRAQLITGATLQDSGRYKCRGAGGACSS  
 IIVRHAREVFEQALKOLEVLEGGAAATLRCVLSVAAPKWCYGNVRLPDGKYSLRQ  
 EGAMLELVNRNLPQDSGRYSCSFGDTTATLTVTALPAQFIGKLRNKEATEGATAT  
 LRCELSTAPVWEKRGSETLRDGRYCLRDQDAMCELCIRGLAMVDAEYSCVCEER  
 TSASLTIRPMAHFIGRLHQSIEGATATLRCELSKAAAPVWEKRGESLVDGRHSL  
 RODGAVCELOICGLAVADAGEYSCVCEERTSATLTVKALPAKTEGLRNEEA VEGAT  
 AMLWCELSKAPVWEKRGPNLRDGRYILRQEGTRCELQICGLADADAGEYLCVCCQ  
 ERTSATLITRALPARFIEDVKNQAREGATAVLOCELSNAPVWRKSGSETLRDGRY  
 SLRQDTCCELQIRGLAMADTGEYSCVCGQERTSMTLTVRALPIKTEGLRNEEAETG  
 ATAVRLCELSMAPEWVKHETLRDGRHSLRQDGRGARCCELQIRGLVAEADAGEYLCM  
 CKERTSMTLVRAMPSKFI EGLRNEEAETGATLWCELSKAAAPVWEKRGHETLRDGR  
 RHLRQDSRCELQIRGLVADAGEYSCVCGQERTSATLTVRALPARFIEDVKNQEAR  
 EGATAVLOCELSKAAAPVWEKRGSETLRDGRYSLRQDTRCELQICGLVADAGEYSC  
 VCGERTSATLTVRALPARFIEDVKNQAREGATAVLOCELSNAPVWRKSGSETLRDGRY  
 ANGRREPLQTAELVLQDREDTGYTCTCGSQATSATLTVTAAPVRLRELQHQ  
 EDEEGTAHLCCELSRAGASVWEKRGSLQFPCKAYQVQDGAALAEILLVGEVEDAG  
 DYTCDTGTQSMASLVRVPRKFTLRQLEQETGDIARLCCOLDAESGAVQWMLK  
 EGVLEHAGPKYEMRSQGATRELLIHLQEAKTGEYACTVGGQTAASLVRTEPVTV  
 RGLVDAETADEVEFCEVSRAGATGVQWCLQPLQSNVETVAVDRGRIHTLRUK  
 CVTPEDATGVFHLGNHASSAQLTVRAPEVTLTLEPLQDVLSEGDASFOCLRBSAG  
 QEARWALGGVLOANEMNDITVEQGTLLHLTLHKVTLLEDAGTVSFHVTGTCSSAQLKV  
 TAKNTVGLNLEVALEGEALFECQLSQPEVAATWLLDDEPVTSENAEAVFENG  
 LRHLLLNLPQDSQCRVTVLAGDMVTSAFTVAGWRLEILEPLKNAARVAGAQT  
 CTLSEAVPGEASWYINGAAVQDDSDMTVADGSHQALLRSGAPHAQETVFPACRD  
 AVASARLTGLPDPDABEVAHSHSTVTLISMAAPMSDGGGLCGVRYRVEKGAATQ  
 WRLCHELVPGCEVVDGLAGETVRFVAAVPGVAGEPVHLPTVLAEPPLPVPVQ  
 PASEQDAGEVDSLELVVAEAGEVIMHKGWERIQPGGRFEWVSQGRQMLVKG  
 FTADQGEYHGLCAQSGICPAAATFQVALSPASVDEAPQSLPEPAQEGEGLLHWEA  
 LARKRREPTLDSISLSELPEDRSQRLQAEVAPDLSSEGTSTADELALGDADL  
 SHTSDDESRAGTSLVTLKAGRPGTSLASKVGAAPASVVKPQDQFLAARPP  
 LGDSTKDGLDSMDKAAVKITQAAFKYKVRKEMKQEGPMFTQDGTFAQVGDALR  
 LECVASKADVRAASLKDGBELTDRGHHIIDLQDGTCSLLIAGLRADAGCTCVQS  
 NKFGVTHSCAVVVSSESESSGELDDAPRAARLRLHRTKSPAEVDEBLF  
 LSADGPAEPEPADMOTYBDEHPI CIRFEALTAEPAQVTRQEMFATIGIVEIKL  
 VEQGRFVEMCISKETAPVVPPEPLSLTSDAAPVFLTEQLQNEQVQDGPVPSFCV  
 VTGPMFPEVRKQKULEDDHYNEDQGGHLLITAVFADMGVYRCLAEANGV  
 BSLTKAELRDLTSDYTDADATESYFSAQYLSRSEQETGETTDBGQPPQVVE  
 VLRDQVAPGLTFLAKFQKYPAPRLYMKDQQLTSAHIMTDKTLTLEIIS  
 VTRDSGGYAAIISNMAAAYSSARLLVRGDEPEEKPSDQVHEQLVPPMLERFTPK  
 KVKSGSTTFEFGVEGPRVPTVHMLREEAERGLWIGDTPGTVASSAQHSLVLLD  
 VROHQGTYTICIASNAAGQALCSASLHVSGLPKVEOEKVEKALISTFLOGTQALISA  
 QGPTASPADLGGQRKEEPLAALHGLSLAEVGTBEFLKLTSLQITTEMASAKITQA  
 KLVQPGSDSDSTKPSAPRHKSRPSSISQSSSESDGDGARGEIDFYVVTADLV  
 PLGAQDAITLREGVVEVLDAAHPLRLVTRFTKSPRSRQSPAYLRKLSPL  
 EWGAAPAEFPEGEAVSEDEYKARLSVYQELLSQAEVFEELQFQSHHQLHRCCHP  
 VPIAQAQVAVIIFRNRDIGRHSFLLQELQDQDTPDDVAMCTIKNOAAEQVLEFLV  
 RQVAVESVSTAIQEFYKVAEALLAGDPSPPPPLQYLEQVQVQRYQVQALIK  
 ELIKARNRPNKCALLBQAVASVALPQRAENKLHSLMENYPCITLQALGEPHQH  
 IHWEGAGARMPKWHNEHVLFRNHLVICKPRDSDRTDTSVYFRNMKLSPISDLND  
 QVEGDDARFVWQREDRSVKLILQARTAIKISWKEICIQRLRLKLVWPPDFFE  
 ELADCTAELGTEVKLACKVTGTPKPIVSWYKDGKAVQVDPHLLIEDPDSGAILDLS  
 LTVGDSQCMCFASAGNCTGKILVQPPFVFNKVASPFVGEADGQFTCTIEGA  
 PYPQIVRQKCALLITGNKFTLSEPSGLLVLVRAASKEDLGLYCELLNRLGASR  
 ASBLRQSQMLQAOCHREQLVAEVTEQETKPKTVIETITVTKVSPRGOR  
 RPSKSPSRSPSCASPLREGLLAPLLYLPAGAGPRRPEABPGKVPVPTLVTEA  
 BAHSPALPLGSPQPKWVEETIEVRKMGQGSPTTEVPRSSGHLFTLPGATP  
 GGPNSNNNNKLLQAQAGTAVGVRPPLVFRVDARSVDAWASGMSLSEEGTM  
 BEAGEEGEDGDAFVTEESQDTHSLGRDPRKILTHNGRMULTLADLEDVPEGEETPHC  
 EGGPGFADPPCEVSIQREIPEPTVYSHVHALGPRGLFGRPEAGSPFP  
 GPQVRSLEGTSFLLREAPARPVGSAWPWTQSFCTIRRSADSGOSSFTTSLSTQVTFNG



ORIGIN	TVGTEVTLLHICPDRDRDDEAAQP"	mRNA	join(<4472..4747,5585..5860,6702..6977,7817..8092,11254..11529,17940..>18215)
Query Match	100.0%; Score 485; DB 9; Length 20435;		/gene="OBSCN"
Best Local Similarity	100.0%; Pred. No.1e-99;	CDS	join(<4472..4747,5585..5860,6702..6977,7817..8092,11254..11529,17940..>18215)
Matches 485;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;		/gene="OBSCN"
			/codon_start=3
QY	1 GACGTGGTACAGAGCGGAGAGAGCTCAGCTCCAGCTCAAAAGTACGATGGAGGTCAA		/product="obscurin"
Db	3652 GACGTGGTACAGAGCGGAGAGAGCTCAGCTCCAGCTCAAAAGTACGATGGAGGTCAA		/protein_id="CAC85747.1"
			/db_xref="GI:21104330"
QY	61 GGGCTGCACAGAAAGCGCTGGTAGTGCAGCAGGTGGGCAAGCAGATGCTGGGGAGTACAG		/db_xref="SPTREMBL:Q8NH7"
Db	3722 GGGCTGCACAGAAAGCGCTGGTAGTGCAGCAGGTGGGCAAGCAGATGCTGGGGAGTACAG		/translation="PKVVFPAKQVAHSEYQAEAGASATLSCEVAQAQTEVMVYKDGKK
QY	121 CTGCGAGCTGGGGCCAGAGAGTCTCTTTCAACTGACATCAAGAGCCCAAGGCGAGT		LSSLVKHVVEAKGCRRLVVOAQGTDAAGDYSCEARGORVSRLLHTEPKVWFKEQOS
Db	3782 CTGCGAGCTGGGGCCAGAGAGTCTCTTTCAACTGACATCAAGAGCCCAAGGCGAGT		VNEVQAEAGAGAMLSCEVAQAQTEVTWKDKKKLSSSKVGMVKGCTRRRLVLPQAG
QY	181 GTTTCGCCAAGAGCAGTGTGGTGCAATAGTGGGACTGAGCGAGCGGCGAGTGGCCAC		KADAGYCEAGGQQRVSFHLHTEPKVGFPAKQSVHNEVQAEAGTAMLSCEVAQFQT
Db	3842 GTTTCGCCAAGAGCAGTGTGGTGCAATAGTGGGACTGAGCGAGCGGCGAGTGGCCAC		EVTVYKDGKKLSSSKVRNEVKGCTRRRLVVOQVKADAGEYSCGAGQQRVSFQHLHTE
QY	241 ACTGAGCTGTGAGTGGGCCCGCCAGACAGAGTGACGTGGTACAGGATGGGAAGAA		PKVFAKEQLVHNEVTEAGASATLSCEVAQAQTEVTWKDKKKLSSSKVRIEAGC
Db	3902 ACTGAGCTGTGAGTGGGCCCGCCAGACAGAGTGACGTGGTACAGGATGGGAAGAA		MQQLVVOAQAGTGTCEAGQRLSFHLDSVEPKVPAKEQLAHRKQVABAGATAT
QY	301 GCTGAGCTCCAGTTCGAAAGTGGCATAGAGGCTCGCGGCTGCATGGCGGAGCTGGTGGT		LSCEVAQAQTEVTWKDKKKLSSSKVRMEAVGCTRRRLVVOACCAQDTEYSCGAGGQ
Db	3962 GCTGAGCTCCAGTTCGAAAGTGGCATAGAGGCTCGCGGCTGCATGGCGGAGCTGGTGGT		RLSFSLDVAEPKVVFPAKEQPVHREVQQAQASTTLLSCEVAQAQTEVMVYKDGKKLSFS
QY	361 GCAGCAGCAGCGCCAGGAGATGCTGGGGAGTACACCTGTGAGGCTGGGGCCAGCGGCT		SKVRFVAVGCTRRRLVVOAQQAQVAGEYSCGAGSQRLLSFHLHVA"
Db	4022 GCAGCAGCAGCGCCAGGAGATGCTGGGGAGTACACCTGTGAGGCTGGGGCCAGCGGCT		<4472..4747
QY	421 CTCCTTCACCTGATGTTTCAGAGCCCAAGCGGCTGTTTCCAAAGAGAGCTGGGACA		/gene="OBSCN"
Db	4082 CTCCTTCACCTGATGTTTCAGAGCCCAAGCGGCTGTTTCCAAAGAGAGCTGGGACA		/note="all"
QY	481 CAGGA 485		4748..5584
Db	4142 CAGGA 4146		/gene="OBSCN"
			/note="all"
RESULT 2			5585..5860
HSA314900	HSA314900 19475 bp DNA linear PRI 21-MAY-2002		/gene="OBSCN"
LOCUS	Homo sapiens partial OBSCN gene for obscurin, exons all-a16.		/note="a11"
DEFINITION	Homo sapiens partial OBSCN gene for obscurin, exons all-a16.		5861..6701
ACCESSION	AJ314900		/gene="OBSCN"
VERSION	AJ314900.1 GI:21104329		/note="a12"
KEYWORDS	OBSCN gene; obscurin.		6702..6977
SOURCE	Homo sapiens (human)		/gene="OBSCN"
ORGANISM	Homo sapiens		/note="a12"
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		6978..7816
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		/gene="OBSCN"
REFERENCE	1		/note="a13"
AUTHORS	Young, P., Ehler, B. and Gautel, M.		7817..8092
TITLE	Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere assembly		/gene="OBSCN"
JOURNAL	Unpublished		/note="a14"
REFERENCE	2 (bases 1 to 19475)		8093..11253
AUTHORS	Gautel, M.S.		/gene="OBSCN"
TITLE	Direct Submission		/note="a14"
JOURNAL	Submitted (22-MAY-2001) Gautel M.S., Physikalische Biochemie, Max-Planck-Institut fuer molekulare Physiologie, Otto-Hahn-Strasse 11, Dortmund, 44227, GERMANY		11254..11529
FEATURES	Location/Qualifiers		/gene="OBSCN"
source	1..19475		/note="a15"
	/organism="Homo sapiens"		17940..>18215
	/mol_type="genomic DNA"		/gene="OBSCN"
gene	/db_xref="taxon:9606"		/note="a16"
	4472..18215		
	/gene="OBSCN"		
Query Match		57.6%; Score 279.4; DB 9; Length 19475;	
Best Local Similarity		99.6%; Pred. No.3.5e-53;	
Matches 280;		Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	163 CACAGAGCCCAAGGCGAGTGTTCGCAAGGAGCAGTTGGTGCAATAGAGTGGGACTGA		222
Db	7812 CTGAGAGCCCAAGGCGAGTGTTCGCAAGGAGCAGTTGGTGCAATAGAGTGGGACTGA		7871
QY	223 GGCAGGGGCGAGTGGCCACACTGAGCTGTGAGGTGGCCAGGCCCAAGAGAGGTGACGCTG		282
Db	7872 GGCAGGGGCGAGTGGCCACACTGAGCTGTGAGGTGGCCAGGCCCAAGAGAGGTGACGCTG		7931
QY	283 GTACAAGGATGGAGAGAACTGAGCTCCAGTTTCGAAAGTGCATAGAGGCTGCGGGCTG		342
Db	7932 GTACAAGGATGGAGAGAACTGAGCTCCAGTTTCGAAAGTGCATAGAGGCTGCGGGCTG		7991
QY	343 CATGCGGCGAGTGGTGGTGAGCAGCAGGCGCCAGGAGATGCTGGGGAGTACACCTGTGA		402



intron	/gene="OBSCN" /number=5 593. .947 /gene="OBSCN" /number=5 948. .1256 /gene="OBSCN" /number=6 1257. .1719 /gene="OBSCN" /number=6 1720. .1986 /gene="OBSCN" /number=7 1987. .2577 /gene="OBSCN" /number=7 2578. .2859 /gene="OBSCN" /number=8 2860. .3169 /gene="OBSCN" /number=8 3170. .3445 /gene="OBSCN" /number=9 3446. .5480 /gene="OBSCN" /number=9 5481. .5750 /gene="OBSCN" /number=10 5751. .10615 /gene="OBSCN" /number=10 10616. .10891 /gene="OBSCN" /number=11
exon	
intron	
exon	
intron	
exon	

```

Query Match      45.5%; Score 220.8; DB 9; Length 24545;
Best Local Similarity 81.7%; Pred. NO. 6.3e-40;
Matches 255; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY      163  CACAGAGCCCAAGGCAGCTGTTTGCACAGAGCAGTGTGTCATATCAGGTGCGGACTGA 222
DB      10611  CCCAGAGCCCAAGGTGTTTGTCTAAGAGCAGCTGGCAGCAGGAGCTTCAGGCAGA 10670

QY      223  GGCAGGGGCAGCTGCCACACTGAGCTGTGAGTGGCCCCAGGCCACAGAGGTGACGTG 282
DB      10671  GGCAGGAGCCAGCTGCCACACTGAGCTGCAGGTGGCCCAAGGCCACAGGAGTGCAGTG 10730

QY      283  GTACAGGATGGAGAGAGCTGAGCTCCAGTTCGAAATGCGCATAGAGGCTCGGGAGTG 342
DB      10731  GTACAGAGTGGAGAGAGCTGAGCTCCAGTTCGAAATGTGCATGAGGGCCACAGGCTG 10790

QY      343  CATCGGCGAGCTGTGGTGTCAGCAGCAGGCGCAGCAGATCTGGGAGTACACCTGTGA 402
DB      10791  CACGCGCAGGCTGTGTTGTGCAGCAGCAGGCGCAGCGGATCCGCGGAGTATAGTCCGA 10850

QY      403  GCGTGGGGCCACGGGCTCTCTCTCCAGCTGGATGCTTCAGAGCCCAAGGCGGTGTTCG 462
DB      10851  GCGTGGGGCCAGCGGCTCTCTCTCCATCTGGATGTCAAAGGTCAAGTGTATAGAGCAGAC 10910

QY      463  AAAGGAGCAGCT 474
DB      10911  ATTTGAGCATCT 10922

RESULT 5
AX883644
LOCUS
DEFINITION Sequence 18549 from Patent EP1074617.
                2488 bp      DNA      linear      PAT 17-DEC-2003

```

AX893644  
VERSION AX893644.1 GI:40038545

SOURCE Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,  
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki.T.  
TITLE Primers for synthesizing full-length cDNA and their use  
JOURNAL Patent: EP 1074617-A 18349 07-FEB-2001;  
Research Association for Biotechnology (JP)

FEATURES Location/Qualifiers  
1..2488 /organism="Homo sapiens"  
         /mol\_type="unassigned DNA"  
         /db\_xref="taxon:9606"  
         697\_...1197  
         /note="unnamed protein product"

CDS  
     /codon\_start=1  
     /protein\_id="CAE92053.1"  
     /gb\_xref="GI:40038546"  
     /translacion="MALHSGMENSCDAGRWMGFSQLHGQCOAWEGSVSPFSSRPSY  
PWFUTLWAVSLTSCVPPCPSEPKVPFAKEQAHPREVOAEAGASATLSCEVAOAOCTE  
PVNYKDGLKLLSSSKVRVEAUGCITRLRVVQQAGQABGEYSCEAGGGQLSLPELVQAGC  
CFGDAE"

ORIGIN  
  
Query Match      45.1%; Score 218.6; DB 6; Length 2488;  
Best Local Similarity     86.1%; Pred. No. 2.8e-39;  
Matches 242; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 163 CACAGAGCCCAAGGCAGTCCTTGCCAAAGGAGCAGTTGTGTCATAATGAGTGGCGACTGA 222  
Db 891 CCAGAGGCCCAAGTCGTTGTTGCCAAGGAGCACGACCACAGGGAGGTGCAGCTGA 950  
Qy 223 GGCGGGGGGCGAGTGCACACTGAGCTGTGAGGTGGCCCCAGGCCCACAGACAGGTACACGTG 282  
Db 951 GGCGGGGGGCGAGTGCACCTGAGCTGCAGTGTGCGAGTGGCCCCAGGCCCACAGACAGGTACG 1010  
Qy 283 GTACAAGATTGGAGAAGACTGAGCTCCAATTGCMAAGTGGCATAGAGCTGCGGGCTG 342  
Db 1011 GTAACAAGATTGGAGAAGACTGAGTTCACAGTTCGAAGTGCAGTGGAGCGCTGGGGCTG 1070  
Qy 343 CATGCGGCGAGCTGGTGGTGCAGCAGGCGAGCCAGGCAGAGTAGTCTGGSGAGTACACCTGTGA 402  
Db 1071 CACACGAGGCTTGGTGGTGCAGCAGCGGGCCAGGCAGAGGCGGGGAGTACAGTGC GA 1130  
Qy 403 GGCTGGGGGCGAGCGGCTCTCCTTCCACTCGGATGTTTCAG 443  
Db 1131 GGCAGGGGTCAGCAGCTCTCCTTCGCGCTGCAGTGGCAG 1171

RESULT 6  
BD160445 2488 bp DNA linear PAT 17-JAN-2003  
LOCUS Primer for synthesizing full-length cDNA and use thereof.  
DEFINITION ACCESSION  
BD160445  
VERSION BD160445.1 GI:27866203  
KEYWORDS JP 2002191363-A/15288.  
SOURCE Homo sapiens (human)  
OS Homo sapiens (human)  
PN JP 2002191363-A/15288

REFERENCE  
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,  
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki.T.  
TITLE Primers for synthesizing full-length cDNA and use thereof  
JOURNAL Patent: JP 2002191363-A 15288 09-JUL-2002;  
HELIIX RESEARCH INSTITUTE  
OS Homo sapiens (human)  
PN JP 2002191363-A/15288

PD 09-JUL-2002  
 PF 28-JUL-2000 JP 2000280990  
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
 PI SAITO,  
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
 PI KEIICHI NAGAI, TETSUJI OTSUKI  
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
 10,  
 PC C12P21/02, C12Q1/68, C12P21/08, C06F17/30, C12N15/00, C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof PH Key

FT CDS Location/Qualifiers  
 (697).. (1194).

## FEATURES

source

1..2488  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

## ORIGIN

Query Match 45.1%; Score 218.6; DB 6; Length 2488;  
 Best Local Similarity 86.1%; Pred. No. 2.8e-39;  
 Matches 242; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
 QY 163 CACAGAGCCCAAGCAGTGTTCACAGGAGCAGTGTGTCATATAGTGGGACTGA 222  
 Db 891 CCCAGAGCCCAAGTGTGTTCACAGGAGCAGCAGCAGGAGGTGCAGGCTGA 950  
 QY 223 GGCAGGGGCCAGTGCACACCTGAGCTGAGTGGCGCCAGCCAGCCAGAGGTGACGTG 282  
 Db 951 GGCAGGGGCCAGTGCACACCTGAGCTGAGTGGCGCCAGCCAGAGGTGACGTG 1010  
 QY 283 GTACAGAGTGGGAAGAGTGCCTCAGTTCGAAAGTGGCGCATAGAGCTCGGGCTG 342  
 Db 1011 GTACAGAGTGGGAAGAGTGCCTCAGTTCGAAAGTGGCGCATAGAGCTCGGGCTG 1070  
 QY 343 CATCGGGCAGCTGGTGGTGCAGGAGGAGCCAGGAGTGGGGAGTACACCTGTGA 402  
 Db 1071 CACAGGAGGCTGGTGGTGCAGGAGGAGCCAGGAGTGGGGAGTACACCTGTGA 1130  
 QY 403 GCCTGGGGCCAGCGGCTCTCTCCACCTGGAGTTTCAG 443  
 Db 1131 GGCAGGGGTCAGCAGCTCTCTCCGCTGCAGGTGGCAG 1171

## RESULT 7

AK024186

LOCUS Homo sapiens cDNA FLJ14124 fis, clone MAMMA1002498. 2488 bp mRNA linear PRI 01-AUG-2002

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,  
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan; cDNA full insert  
 sequencing; Research Association for Biotechnology; cDNA library

construction, 5'- & 3'-end one pass sequencing and clone selection:  
 Helix Research Institute (supported by Japan Key Technology Center  
 etc.) and Department of Virology, Institute of Medical Science,  
 University of Tokyo.

## FEATURES

source

Location/Qualifiers  
 1..2488

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MAMMA1002498"

/tissue\_type="Mammary gland"

/clone\_lib="MAMMA1"

/note="cloning vector: pME18SFL3"

697..1197

/note="unnamed protein product"

/codon\_start=1

/protein\_id="BAB14847.1"

/db\_xref="GI:10436504"

/translation="MALHSGWRNCDSCGRWGFSLHGCQANEGSVSPTFSRPSY  
 PRVFLTWAVSLTSCVPPSPKPVFAKEQPAHREVQAEAGASATLSCEVAQAQTE  
 VTYKDGKLSKSSKVRVEAVGCTRLRVQOAGQAEAGSVCEAGGQQLSRLQVAGQ  
 CFGDAE"

## ORIGIN

Query Match 45.1%; Score 218.6; DB 9; Length 2488;  
 Best Local Similarity 86.1%; Pred. No. 2.8e-39;  
 Matches 242; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
 QY 163 CACAGAGCCCAAGCAGTGTTCACAGGAGCAGTGTGTCATATAGTGGGACTGA 222  
 Db 891 CCCAGAGCCCAAGTGTGTTCACAGGAGCAGCAGCAGGAGGTGCAGGCTGA 950  
 QY 223 GGCAGGGGCCAGTGCACACCTGAGCTGAGTGGCGCCAGCCAGAGGTGACGTG 282  
 Db 951 GGCAGGGGCCAGTGCACACCTGAGCTGAGTGGCGCCAGCCAGAGGTGACGTG 1010  
 QY 283 GTACAGAGTGGGAAGAGTGCCTCAGTTCGAAAGTGGCGCATAGAGCTCGGGCTG 342  
 Db 1011 GTACAGAGTGGGAAGAGTGCCTCAGTTCGAAAGTGGCGCATAGAGCTCGGGCTG 1070  
 QY 343 CATCGGGCAGCTGGTGGTGCAGGAGGAGCCAGGAGTGGGGAGTACACCTGTGA 402  
 Db 1071 CACAGGAGGCTGGTGGTGCAGGAGGAGCCAGGAGTGGGGAGTACACCTGTGA 1130  
 QY 403 GCCTGGGGCCAGCGGCTCTCTCCACCTGGAGTTTCAG 443  
 Db 1131 GGCAGGGGTCAGCAGCTCTCTCCGCTGCAGGTGGCAG 1171

## RESULT 8

AC023889

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (18-FEB-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 On Jul 7, 2000 this sequence version replaced gi:8748947.



CDS  
 join(<1265..1540,5880..>6155)  
 /gene="OBSCN"  
 /codon\_start=3  
 /product="obscurin"  
 /protein\_id="CAC85749.1"  
 /db\_xref="GI:21104332"  
 /db\_xref="SPTREMBL:Q8NH6"  
 /translation="PKAVFAKEQPAHREVOAEAGASATLSCEVAQAQTEVTWYKDGKK  
 LSSSLKVEAGTRRLVVOQAGADTGYSCAQQQLSFLQVAEPRAVFAKEQP  
 ASREVOAAGTSATLSCEVAQAQTEVTWYKDGKKLSSSKVRMEAVGCTRRLLVVQEAG  
 QADAGEVSKAGDQRLSFLHVA"  
 <1265..1540  
 /gene="OBSCN"  
 /note="B"  
 1541..5879  
 /gene="OBSCN"  
 /number=16  
 5880..>6155  
 /gene="OBSCN"  
 /number=17

exon  
 intron  
 exon

ORIGIN  
 Query Match 44.5%; Score 216; DB 9; Length 10091;  
 Best Local Similarity 80.8%; Pred. No. 8.7e-39;  
 Matches 252; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
 QY 155 CTGCACATCAGAGCCCAAGCGAGTGTTCCTCCAGAGAGAGTGTGTCATATAGAGTG 214  
 Db 1252 CTGGCTTCCAGAGCCCAAGCGAGTGTTCCTCCAGAGAGAGTGTGTCATATAGAGTG 1311  
 QY 215 CGGACTGAGGAGGGGGCCAGTGCACACTGAGCTGTGAGGTGGCCAGGCCACAGACAG 274  
 Db 1312 CAGGCTGAGAGAGGGGCTAGCGCACACTGAGCTGTGAGGTGGCCAGGCCACAGACAG 1371  
 QY 275 GTGACGTGGTACAGAGTGGAGAGTGTGCTCCAGTGTGCTGAAAGTGGCGATAGAGGCT 334  
 Db 1372 GTGACTTGGTACAGAGAGGGGAGTGTGCTCCAGTGTGCTGAAAGTGGCGATAGAGGCT 1431  
 QY 335 CGGGGCTGCTGCGGCGAGCTGGTGGTGCAGAGAGCCAGGCGAGTGTGCGGGAGTAC 394  
 Db 1432 CGAGCTGTACAGAGAGCTGGTGGTGCAGAGAGCCAGGCGAGTGTGCGGGAGTAT 1491  
 QY 395 ACCTGTAGGCTGGGGCCAGCGCTCTCCCTTCACCTGGATGTTTCAGAGCCCAAGGCG 454  
 Db 1492 AGCTGTAGGCTGGGGCCAGCGCTCTCCCTTCGCGCTGCAAGTGGCAGGTCAAGTGT 1551  
 QY 455 GTGTTTCCAAAG 466  
 Db 1552 GGGGATGCTGAG 1563

RESULT 10  
 AK128447/c  
 LOCUS Homo sapiens cDNA FLJ46590 fis, clone THYMU304441. PRI 09-SEP-2003  
 DEFINITION  
 ACCESSION AK128447  
 VERSION AK128447.1 GI:34535823  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1  
 Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T.,  
 Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K.,  
 Kamiyama, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M.,  
 Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A.,  
 Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K.,  
 Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y.,  
 Sekine, M., Kikuchi, H., Murakawa, K., Kanehori, K.,  
 Takahashi-Fujii, A., Ohshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,  
 Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.  
 NEDO human cDNA sequencing project

TITLE

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3956)  
 AUTHORS Isogai, T. and Yamamoto, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan  
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
 Research Association for Biotechnology (RAB); cDNA library  
 construction: Helix Research Institute (HRI) (supported by Japan  
 Key Technology Center etc.); 5' and 3' end one pass sequencing: RAB,  
 HRI, and Biotechnology Center, National Institute of Technology and  
 Evaluation; clone selection for full insert sequencing: HRI and  
 RAB; annotation: HRI and RAB.

FEATURES  
 source  
 1..3956  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="THYMU304441"  
 /tissue\_type="thymus"  
 /clone\_lib="THYMU3"  
 /note="cloning vector: pME18SFL3"  
 1014..1415  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAC87445.1"  
 /db\_xref="GI:34535824"  
 /translation="WSPPSGPGPHSSAWHPPPPPAPPWCPAAPPWQTPPWALGWTRR  
 DAGSEAPQTQPMQCPQSDSQAGLSIIKQCTVTHRAPPTTRAKGLRLHVGSTIISLG  
 NRMRLPCLGLPFGWPHLSPALPTTQGRQ"

ORIGIN  
 Query Match 43.5%; Score 211.2; DB 9; Length 3956;  
 Best Local Similarity 83.3%; Pred. No. 1.2e-37;  
 Matches 240; Conservative 0; Mismatches 48; Indels 0; Gaps 0;  
 QY 156 TGCACATCACAGAGCCCAAGCGAGTGTTCCTCCAGAGAGAGTGTGTCATATAGAGTGC 215  
 Db 1139 TGTCCATCCAGAGCCCAAGGTGGTGTTCCTCCAGAGAGAGTGTGTCATATAGAGTGC 1080  
 QY 216 GGACTGAGCGAGGGGCCAGTGCACACTGAGCTGTGAGGTGGCCAGGCCACAGACAGG 275  
 Db 1079 AGGCGAGGCGGGGCCAGTGCACACTGAGCTGTGAGGTGGCCAGGCCACAGATGGAGG 1020  
 QY 276 TGACGTGTACAGAGTGGGAAGAGCTGAGCTCCAGTTCGAAAGTGGCATAGAGGCTG 335  
 Db 1019 TGACATGTGTACAGAGCGGGAAGAGTGTGAGCTCAAAAGTGGCATAGAGGCCA 960  
 QY 336 CGGGCTGCATGCGGCGAGCTGGTGTGTCAGAGAGGCGCCAGGCGAGATGCTGGGAGTACA 395  
 Db 959 GCGGCTACACAGCGAGGCTGGTGTGTCAGAGAGGCGGCGGCGAGATGCTGGGAGTACA 900  
 QY 396 CTTGTGAGGCTGGGGGCCAGCGGCTCTCTTCACCTGGATGTTTCAG 443  
 Db 899 GCTGTGAGGCGGGGGCCAAACGGTGTCTCTTCGCGCTGCACGTGGCGAG 852

RESULT 11  
 HSA314903  
 LOCUS Homo sapiens partial OBSCN gene for obscurin, exons C-22.  
 DEFINITION  
 ACCESSION AJ314903  
 VERSION AJ314903.1 GI:21104333  
 KEYWORDS OBSCN gene; obscurin.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1  
 Young, P., Ehler, E. and Gautel, M.  
 Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere

Query Match 43.2%; Score 209.6; DB 9; Length 6400;  
Best Local Similarity 83.0%; Pred. No. 2.6e-37;

Query Match 42.8%; Score 207.6; DB 10; Length 211829;  
 Best Local Similarity 82.9%; Pred. No. 4.4e-37;  
 Matches 237; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 165 CAGAGCCCAAGCAGTGTTCGCAAGGACGAGTGTGTCATATAGGTGCGGACTGAGG 224  
 Db 91363 CAGAGCCCAAGCAGTGTTCGCAAGGACGAGTGTGTCATATAGGTGCGGACTGAGG 91304

QY 225 CAGGCGCAGTGCACACTGAGTGTGAGTGGCCAGCCAGCCAGACAGAGTGCAGTGGT 284  
 Db 91303 CTGGGGCCAGTGCACACTGAGTGTGAGTGGCCAGCCAGCCAGACAGTGCAGTGGT 91244

QY 285 ACAAGGATGGGAAGAGTGCAGTCCAGTTCGAAAGTGCATAGAGCTCGGGCTGCA 344  
 Db 91243 TCAAGGCGGGAAGAGTGCAGTCCAGTTCGAAAGTGCATAGAGCTCGGGCTGCT 91184

QY 345 TCGGCGCAGTGTGTCAGCAGGAGCCAGCCAGGATGCTGGGAGTACACCTGTGAGG 404  
 Db 91183 CAGGAGGCTGTGTCAGCAGGAGCCAGCCAGGATGCTGGGAGTACACCTGTGAGG 91124

QY 405 CTGGGGGCGCAGCGGCTCTCTCTTCCACCTGGGATGTTTCAGAGCCCAA 450  
 Db 91123 CGGGGGCCAGAGTCTCTCTTCCACCTGGGATGTTTCAGAGCCCAA 91078

RESULT 13  
 AC099089/c  
 LOCUS Rattus norvegicus clone CH230-154E3, WORKING DRAFT SEQUENCE.  
 AC099089  
 AC099089 5 GI:30522701  
 HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
 KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE Rattus norvegicus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 260998)  
 Murny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,  
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
 Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,  
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
 Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
 Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,  
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
 Jackson,B., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kowis,C., Kraft,C.B., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorensheva,L., Louisseg,H., Lozada,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
 Nankervis,C., Neal,D., Newton,S., Nguyen,N., Norris,S.,  
 Nwackeleneh,O., Okwunonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,  
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,

Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trujos,Z., Umani,K.,  
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstein,G. and Gibbs,R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 260998)  
 Worley,K.C.  
 Direct Submission  
 Submitted (09-NOV-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 260998)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On May 10, 2003 this sequence version replaced gi:22855456.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GGLZ  
 Center clone name: CH230-154E3  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0;  
 Consensus quality: 216929 bases at least Q40  
 Consensus quality: 219228 bases at least Q30  
 Consensus quality: 220845 bases at least Q20  
 Estimated insert size: 226118; sum-of-contigs estimation  
 Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 consists of 1 contigs. Gaps between the contigs  
 are represented as runs of N. The order of the pieces  
 is believed to be correct as given, however the sizes  
 of the gaps between them are based on estimates that have  
 been provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 260998: contig of 260998 bp in length.  
 Location/Qualifiers  
 1. 260998  
 /organism="Rattus norvegicus"

FEATURES  
source



```
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-154E3"
misc_feature 1..1486
note="wgs_contig"

ORIGIN
Query Match 42.8%; Score 207.4; DB 2; Length 260998;
Best Local Similarity 83.6%; Pred. No. 4.8e-37;
Matches 235; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 163 CACAGAGCCCAAGCAGTGTTTGCAGAGGACAGTGTGTCATATGAGGTGCGGACTGA 222
DB 188410 CCCAGAGCCCAAGTGTGTTTGCAGAGGACAGTGTGTCATATGAGGTGCGGACTGA 188351
QY 223 GCGAGGGCCCAAGTGTGTTTGCAGAGGACAGTGTGTCATATGAGGTGCGGACTGA 282
DB 188350 GCGAGGGCCCAAGTGTGTTTGCAGAGGACAGTGTGTCATATGAGGTGCGGACTGA 188291
QY 293 GTACAGAGTGGGAGAGCTGAGTCCAGTTCGAGTGGCATAGAGGCTGCGGCTG 342
DB 188290 GTTCAAGGACCGGAGAGCTGAGTCCAGTTCGAGTGGCATAGAGGCTGCGGCTG 188231
QY 343 CATGCGGAGCTGTGTTTGCAGAGGACAGTGTGTCATATGAGGTGCGGACTGA 402
DB 188230 CTCAGAGGAGCTGTGTTTGCAGAGGACAGTGTGTCATATGAGGTGCGGACTGA 188171
QY 403 GCGTGGGGCCAGCGGCTCTCTCCACTCGATGTTTCAG 443
DB 188170 GCGCGGGGACAGAGGCTCTCTCCACTCGATGTTTCAG 188130

RESULT 14
AC026657 164766 bp DNA linear HTG 01-SEP-2000
LOCUS Homo sapiens chromosome 1 clone AP11-245P10, WORKING DRAFT
DEFINITION AC026657
ACCESSION AC026657
VERSION AC026657.4 GI:9958202
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 164766)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 164766)
Waterston,R.H.
Direct Submission
Submitted (22-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:7637349.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0245P10
----- Summary Statistics -----
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator ET; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 139864 bases at least Q40
Consensus quality: 147866 bases at least Q30
Consensus quality: 151469 bases at least Q20
Insert size: 169000; agarose-fp
Insert size: 161074; sum-of-contigs
```

FEATURES

Location/Qualifiers

Quality coverage: 3.60 in Q20 bases; agarose-fp  
Quality coverage: 3.92 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 31 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1  
\* 1124: contig of 1124 bp in length  
\* 1125: gap of unknown length  
\* 1224: contig of 1808 bp in length  
\* 1225: gap of unknown length  
\* 3032: contig of 1361 bp in length  
\* 3132: gap of unknown length  
\* 3133: contig of 1267 bp in length  
\* 4494: gap of unknown length  
\* 4594: gap of unknown length  
\* 5861: gap of unknown length  
\* 5961: contig of 1710 bp in length  
\* 7671: gap of unknown length  
\* 7771: contig of 1899 bp in length  
\* 9670: gap of unknown length  
\* 9770: contig of 1946 bp in length  
\* 11715: gap of unknown length  
\* 11815: contig of 2428 bp in length  
\* 14243: gap of unknown length  
\* 14343: contig of 2544 bp in length  
\* 16887: gap of unknown length  
\* 16987: gap of unknown length  
\* 19247: contig of 2260 bp in length  
\* 19347: gap of unknown length  
\* 21375: contig of 2028 bp in length  
\* 21476: gap of unknown length  
\* 25026: contig of 3550 bp in length  
\* 25126: gap of unknown length  
\* 28051: contig of 2926 bp in length  
\* 28151: gap of unknown length  
\* 32054: contig of 3903 bp in length  
\* 32154: gap of unknown length  
\* 35716: contig of 3562 bp in length  
\* 35816: gap of unknown length  
\* 35717: contig of 5075 bp in length  
\* 35817: gap of unknown length  
\* 40892: contig of 3036 bp in length  
\* 40992: gap of unknown length  
\* 44127: gap of unknown length  
\* 44128: contig of 4772 bp in length  
\* 48900: gap of unknown length  
\* 49000: contig of 5256 bp in length  
\* 54256: gap of unknown length  
\* 54356: contig of 4639 bp in length  
\* 58994: gap of unknown length  
\* 59094: contig of 5296 bp in length  
\* 59095: gap of unknown length  
\* 64390: contig of 6375 bp in length  
\* 70865: gap of unknown length  
\* 70965: contig of 7702 bp in length  
\* 78667: gap of unknown length  
\* 78668: contig of 8440 bp in length  
\* 87207: gap of unknown length  
\* 87208: contig of 10551 bp in length  
\* 97859: gap of unknown length  
\* 97959: contig of 11562 bp in length  
\* 109520: gap of unknown length  
\* 109521: contig of 11288 bp in length  
\* 120908: gap of unknown length  
\* 121008: contig of 20469 bp in length  
\* 141477: gap of unknown length  
\* 141478: contig of 21095 bp in length  
\* 162672: gap of unknown length  
\* 162673: contig of 1202 bp in length  
\* 163974: gap of unknown length  
\* 163975: contig of 892 bp in length.  
\* 164075: gap of unknown length  
\* 164076: contig of 892 bp in length.

## ORIGIN

Query Match	Score 201.2;	DB 2;	Length 164766;
Best Local Similarity	41.5%;		
Matches	84.1%;		
Mismatches	0;		
Conservative	0;		
Indels	0;		
Gaps	0;		



10/045,072.  
1533.079002

412

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:07:04 ; Search time 1122.53 Seconds  
(without alignments)  
10242.013 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_10286\_10670

Perfect score: 385

Sequence: 1 acctagcctgagcagacg.....cctcagccacgctcaccatc 385

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55025578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_htc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_htc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rod.\*

26: em\_gss\_phg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	285.8	74.2	595	10	BE086827
2	263.4	68.4	486	12	BM030364
3	252.6	65.6	536	10	BF824937
4	240	62.3	723	13	BQ446463

C 5	239.4	62.2	487	10	BE466441
C 6	189	49.1	570	29	CE836029
C 7	186.2	48.4	437	10	BF826423
C 8	175.2	45.5	359	10	BF853346
C 9	168.8	43.8	369	10	BF826489
C 10	165.4	43.0	324	13	BQ361458
C 11	163.2	42.4	601	10	BF398067
C 12	158.8	41.2	690	28	BH025901
C 13	127.2	33.0	505	29	CE078765
C 14	122.4	31.8	204	12	BG987697
C 15	121.8	31.6	647	28	AZ590254
C 16	112.4	29.2	247	12	BG987696
C 17	107.8	28.0	628	28	AZ985514
C 18	106.2	27.6	657	28	AZ508477
C 19	103	26.8	350	10	BF748073
C 20	86.2	22.4	742	13	BH358283
C 21	84.8	22.0	380	10	BF957683
C 22	82.2	21.4	382	9	AI478779
C 23	82.2	21.4	740	13	BH689712
C 24	78.2	20.3	212	10	BF827521
C 25	70	18.2	510	13	EX478907
C 26	69	17.9	330	28	AQ237886
C 27	64	16.6	173	10	BE086762
C 28	57.8	15.0	586	29	AX153687
C 29	54	14.0	332	28	AZ754571
C 30	52	13.5	128	12	BM031705
C 31	52	13.5	997	29	CNS05GFS
C 32	50.4	13.1	722	13	BH322682
C 33	49.4	12.8	1010	29	CNS02HPP
C 34	48.4	12.6	1036	29	CNS00DAP
C 35	47	12.2	1173	28	CC218835
C 36	46.6	12.1	1000	13	EX407619
C 37	46.6	12.1	1970	29	CG748837
C 38	45.8	11.9	828	13	EX425797
C 39	45.6	11.8	855	28	BZ637045
C 40	45.2	11.7	354	10	BF353892
C 41	45	11.7	871	29	CNS04FAX
C 42	44.8	11.6	658	13	EX087878
C 43	44.6	11.6	1181	12	BH69996
C 44	44.4	11.5	1101	29	CNS0153P
C 45	44.4	11.5	1101	29	CNS016JS

## ALIGNMENTS

RESULT 1

BE086827

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

COMMENT

BE086827 595 bp mRNA linear EST 12-JUN-2000  
QV1-BT0678-300400-182-f06 BT0678 Homo sapiens cDNA, mRNA sequence.

BE086827.1 GI:8477221

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

Dias Neto E., Garcia Correa R., Verjovski-Almeida S., Briones M.R.,

Nagai M.A., da Silva W. Jr., Zago M.A., Bordin S., Costa F.F.,

Goldman G.H., Carvalho A.F., Matsukuma A., Baia G.S., Simpson D.H.,

Brunstein A., de Oliveira P.S., Bucher P., Jongeneel C.V.,

O'Hare M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and

Simpson A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202683

10737800

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-Sp,

BE466441 hz21h03.x  
CE836029 tigr-gss-  
BF826423 CM4-HN002  
BF853346 MR3-EN009  
BF826489 CM4-HN002  
BQ361458 PMO-OT023  
BF398067 UI-R-B52-  
BH025901 RPT-24-2  
CE078765 tigr-gss-  
BG987697 PM3-HT116  
AZ590254 1M0399B18  
BG987696 PM3-HT116  
AZ985514 2M0267G22  
AZ508477 1M0347H09  
BF748073 MR2-BN038  
BH358283 603477905  
BF957683 PM1-NN120  
AI478779 tm24e02.x  
BU689712 UI-CF-FN0  
BF827521 CM4-HN002  
EX478907 DKZP686N  
AQ237886 RPT111-68  
BE086762 QV1-BT067  
BX153687 Danilo rer  
AZ754571 cq01e04.f  
BM031705 497184 MA  
AL336218 Tetraodon  
BU322682 603851347  
AL197926 Tetraodon  
AL067111 Drosophil  
CC218835 CH261-69E  
BX407619 BX407619  
CG748837 P042-4-E0  
BX425797 BX425797  
BZ637045 OGCAN08TC  
BF353892 IL5-HT070  
AL288114 Tetraodon  
BX087878 BX087878  
BH69996 603394150  
AL104949 Drosophil  
AL106634 Drosophil

## Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: [animpson@ludwig.org.br](mailto:animpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV1-BF0678-300>)

400-182-106843-2000-04-30&amp;tl=1

Seq primer: puc 18 forward

High quality sequence start: 11

High quality sequence stop: 595.

## FEATURES

source

1..595

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/dev\_stage="Adult"

/clone\_lib="BF0678"

/note="Organ: breast; Vector: puc18; Site: 1: Smai; Site 2:

Smai; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters Patent application No.

196,716 - Ludwig Institute for Cancer Research) profiles

into the pUC 18 vector. Reverse transcription of tissue

mRNA and cDNA amplification were performed under low

stringency conditions."

## ORIGIN

Query Match 74.2%; Score 285.8; DB 10; Length 595;

Best Local Similarity 83.9%; Pred. No. 3.5e-52;

Matches 323; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1 ACATAGCTGAGGACAGCGGGCTGTGTGCGAGCTCAGATCTGTGGCTGTGGCTGTGGC 60

Db 106 ACACAGCCTGAGGACAGCGGGCTCAGGTGTGAGCTCAGATCTGTGGCTGTGGCTGTGGT 165

QY 61 AGATGCTGGGAGTACTCTGTGTGTGGGAGGAGGACCTCTGCCACTCTCACCGT 120

Db 166 GGATGCGGGGAGTACTCTGTGTGTGGGAGGAGGACCTCTGCCACTCTCACCTGT 225

QY 121 GAAGGCGCTGCCAGCCCAAGTTTCACAGAGGTTCTGAGGAATGAAGAGCCGTTGAAGGGGC 180

Db 226 CAGGCGCTGCTCCAGATTTCAGAGATGTGAAAAACAGAGGCCACAGAGAGGGGC 285

QY 181 CACAGCATGTTGTGTGTGTAACCTGACCAAGTGGCCCCCTGTGGAGTGGAGGAGGGCC 240

Db 286 CACGCGCTGTGCAATGTGAGCTGAGCAAGGCGGCCCTGTGGAGTGGAGGAGGGGC 345

QY 241 CGAGAACTCAGAGATGGGACAGATACATCCTGAGGACAGGGGACAGGTGTGAGCT 300

Db 346 TGAGACCTCAGAGGTGGGACAGATACAGCTGAGGACAGTGGACAGATGTGAGCT 405

QY 301 GCAGATCTGGCTGGCCATGGCGACGCGGGGAGTACTTGTGTGTGGCGGAGGA 360

Db 406 GCAGATTATGGCTGTCTGTGGCAGACACTGGGGAGTACTGTGTGTGTGGCGGAGGA 465

QY 361 GAGGACCTCAGCCACGCTCACCATC 385

Db 466 GAGGACCTCGCCACATCACCGTC 490

## RESULT 2

BM030364

LOCUS

DEFINITION 48988 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION

BM030364

VERSION

BM030364.1

KEYWORDS

EST.

ORGANISM

Bos taurus (cow)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

REFERENCE

1 (Bases 1 to 486)

## AUTHORS

Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.M., Rohrer, G.A., Chitko-McKown, C.G., Pettea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and Keefe, J.W.

Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

Genome Res. 11 (4), 626-630 (2001)

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: [smith@email.marc.usda.gov](mailto:smith@email.marc.usda.gov)

Single pass sequencing. Bases called and alt trimmed with phred

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCACGAG

Plate: 120 row: E column: 13

Seq primer: ATTAGTGCACATATAG.

Location/Qualifiers

1..486

/organism="Bos taurus"

/mol\_type="mRNA"

/db\_xref="taxon:9913"

/tissue\_type="pooled"

/lab\_host="DH10B"

/clone\_lib="MARC 2BOV"

/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;

Library made from pooled tissue from testis, thymus,

semitendinosus muscle, longissimus muscle, pancreas,

adrenal, and endometrium."

## ORIGIN

Query Match 68.4%; Score 263.4; DB 12; Length 486;

Best Local Similarity 81.2%; Pred. No. 2.5e-47;

Matches 306; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 9 TGAGCAGAGCGGGCTGTGTGCGAGCTGCAGATCTGTGGCTGTGGCTGTGCAGATCTG 68

Db 73 TGAGCAGAGCGGGCTGTGTGAGCTGGAAATCCATGACCTGACCATGGAGATCTG 132

QY 69 GGGAGTACTCTGTGTGTGTGGGAGGAGGAGGACCTCTGCCACTCTCACCGTGAAGGCC 128

Db 133 GGGAGTACTCTATGCATATGCGGCGAGGAGAAAACCTCAGCCACGCTCGCGCTCAGGGCCC 192

QY 129 TGCCAGCAAGTTCACAGAGGCTGTGAGGATGAAGAGCCGTGGAAGGGGCCACAGCCA 188

Db 193 TGCCCTGCCAAGTTCACAAAAGTCTGAAGAAGGAGAGGCCACAGAAGGGGCCACAGCCA 252

QY 189 TGTGTGTGTGTAACCTGAGCAAGGTGGCCCTGTGGAGTGGAGGAAGGGGCCCGAGAAC 248

Db 253 TACTGCGCTGTGAGCTGAGCAAGCGGCCCTCGTGGAGTGGAGGAAGGGGCCCGAGAAC 312

QY 249 TCAGAGATGGGACAGATACATCTTGGCAGGAGGGGACCCAGGTGTGAGTGCAGATCT 308

Db 313 TCAGAGCGGGGACAGAGTGGCGGTGAAGCAGGAGGGAGCGGTGTGTGAGTGGAGATCC 372

QY 309 GTGGCTCGCCATGGCGAGCCCGGGAGTACTTGTGTGTGTGGGGCAGGAGGAGACCT 368

Db 373 GTGGCTCGCCATGGCGAGTCTGGGAGTACTCTGTCATGTGCGGCGAGGAGACCT 432

QY 369 CAGCCACGCTCACCATC 385

Db 433 CAGCCACAGTCGCGGTC 449

## RESULT 3

BF824937

LOCUS BF824937 536 bp mRNA linear EST 13-JAN-2001  
DEFINITION ILO-HN0039-101100-500-c06 HN0039 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF824937  
VERSION BF824937.1 GI:12167005  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS Dias Neto,B., Garcia Correa,R., Vertovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.P.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,  
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
PUBMED 10737800  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL06t2-IL0-HN0039-  
101100-500-c06&t3=2000-11-10&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 535.  
Location/Qualifiers  
1. 536  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="HN0039"  
/note="Organ: head normal; Vector: puc18; Site:1: Smal;  
Site 2: Smal; A mini-library was made by cloning products  
derived from ORSTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

FEATURES  
source  
1. 536  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="HN0039"  
/note="Organ: head normal; Vector: puc18; Site:1: Smal;  
Site 2: Smal; A mini-library was made by cloning products  
derived from ORSTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

ORIGIN  
Query Match 65.6%; Score 252.6; DB 10; Length 536;  
Best Local Similarity 80.4%; Pred. No. 5.9e-45;  
Matches 308; Conservative 0; Mismatches 74; Indels 1; Gaps 1;  
3 ATAGCCTGAGGAGGAGGCGGTGTGTGCGAGCTCAGATCTGTGGCTGTGGCTGTGGCAG 62  
118 ACAGCCTGAAGCAGGATGGACCACTGTGTGAGCTCAGATCTGTGGCTGTGGCTGTGGCAG 177  
63 ATGCTGGGAGTACTCTCTGTGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 122  
178 ATGCTGGGAGTACTCTCTGTGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 237  
123 AGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 182  
238 GGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 297  
183 CAGCCATGT 242  
298 CAGCTGT 357  
243 AGACCTCAGATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 302

Db 358 AGACCTCAGATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 417  
Qy 303 AGATCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGT 362  
Db 418 AGATCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGT 476  
Qy 363 GGACCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 385  
Db 477 GAACCTCAGCTACACTCACTGTC 499  
RESULT 4  
BQ446463/c  
LOCUS BQ446463  
DEFINITION UI-H-EUI-azz-h-22-0-UI.s1 NCI\_CGAP\_Ccl Homo sapiens cDNA clone  
UI-H-EUI-azz-h-22-0-UI 3', mRNA sequence.  
ACCESSION BQ446463  
VERSION BQ446463.1 GI:21249575  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS 1 (bases 1 to 723)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgaps@mail.nih.gov  
Tissue Procurement: Dr. Jose Mercuende  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bento-soares@uiowa.edu  
Seq primer: M13 FORWARD  
POLYA=Yes.  
Location/Qualifiers  
1. 723  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-H-EUI-azz-h-22-0-UI"  
/tissue\_type="Osteoarthritic Cartilage"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI\_CGAP\_Ccl"  
/note="Organ: Knee; Vector: p773-Pac (Pharmacia) with a  
modified polylinker; Site:1: EcoR I; Site:2: Not I;  
NCI\_CGAP\_Ccl is a normalized cDNA library containing the  
following tissue(s): Osteoarthritic Cartilage The library  
was constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an EcoR I  
adaptor, digested with Not I, and cloned directionally  
into p773-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dfr)18 tail. The sequence tag for this library is  
TGATCAGCT.  
TAG TISSUE=osteoarthritic cartilage  
TAG LIB=UI-H-EUI  
TAG\_SEQ=TGATCAGCT"

ORIGIN  
Query Match 62.3%; Score 240; DB 13; Length 723;  
Best Local Similarity 76.4%; Pred. No. 3.8e-42;  
Matches 294; Conservative 0; Mismatches 91; Indels 0; Gaps 0;  
Qy 1 ACATAGCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 60







```

FEATURES
  source
    Location/Qualifiers
      1..359
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /dev_stage="Adult"
        /clone_lib="EN0091"
        /note="Organ: lung normal; Vector: puc18; Site 1: SmaI;
        Site 2: SmaI; A mini-library was made by cloning products
        derived from ORESTES PCR (U.S. Letters Patent application
        No. 196,716 - Ludwig Institute for Cancer Research)
        profiles into the pUC 18 vector. Reverse transcription of
        tissue mRNA and cDNA amplification were performed under
        low stringency conditions."
ORIGIN
  Query Match      45.5%; Score 175.2; DB 10; Length 359;
  Best Local Similarity 83.5%; Pred. No. 3.9e-28;
  Matches 198; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
  QY 149 GGTCTGAGGATGAAGAGCCGCTGAAGGGGCCACAGCCATGTTGTGTGAACTGAGC 208
  Db 359 GATGTGAGAAATACAGAGGCCACAGAGGGGCCACAGCTGTGCTGCAGTGTGAGCTGAGC 300
  QY 209 AAGTGGGCCCCCTGTGAGTGGAGAGGGGCCCGAGACCTCAGAGATGGGGCAGATAC 268
  Db 299 AAGNCGGCCCCGTGAGTGGAGAGGGGCTGTGAGACCTCAGAGATGGGGCAGATAT 240
  QY 269 ATCTTGAGGAGGAGGACAGGCTGTGAGTGCAGATCTGTGGCCCTGGCCATGGCGGAC 328
  Db 239 AGCCTGAGGAGGACGAGGAGGCTGTGAGTGCAGATCTGTGGCCCTGGCTGTGGAGGAC 180
  QY 329 GCGGGGAGTACTGTGTGTGCGGCGGAGAGGACCTCAGCAGCTCACCATC 385
  Db 179 ACTGAGAGTATTGTGTGTGCGGCGGAGAGAACCTCAGTACACTCACTGTGTC 123

RESULT 9
BF826489
LOCUS
  CM4-HN0020-221100-452-b09 HN0020 Homo sapiens cDNA, mRNA sequence.
  EST 13-JAN-2001
  BF826489
  BF826489.1 GI:12169746
  EST.
  KEYWORDS
  SOURCE
  ORGANISM
    Homo sapiens (human)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 369)
  Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
  Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
  Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
  Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
  O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
  Simpson,A.J.J.
  Shotgun sequencing of the human transcriptome with ORF expressed
  sequence tags
  Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  20202663
  10737800
  Contact: Simpson A.J.G.
  Laboratory of Cancer Genetics
  Ludwig Institute for Cancer Research
  Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
  Brazil
  Tel: +55-11-2704922
  Fax: +55-11-2707001
  Email: asimpson@ludwig.org.br
  This sequence was derived from the FAPESP/LICR Human Cancer Genome
  Project. This entry can be seen in the following URL
  (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM4&t2=CM4-HN0020-
  221100-452-b09&t3=2000-11-22&t4=1)
  Seq primer: puc 18 forward

FEATURES
  source
    Location/Qualifiers
      1..369
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /dev_stage="Adult"
        /clone_lib="HN0020"
        /note="Organ: head normal; Vector: puc18; Site 1: SmaI;
        Site 2: SmaI; A mini-library was made by cloning products
        derived from ORESTES PCR (U.S. Letters Patent application
        No. 196,716 - Ludwig Institute for Cancer Research)
        profiles into the pUC 18 vector. Reverse transcription of
        tissue mRNA and cDNA amplification were performed under
        low stringency conditions."
ORIGIN
  Query Match      43.8%; Score 168.8; DB 10; Length 369;
  Best Local Similarity 78.1%; Pred. No. 9.8e-27;
  Matches 203; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
  QY 3 ATAGCTGAGGAGGAGGCGGCTGTGCGAGCTCAGATCTGTGGCTGGCTGGTGGGAG 62
  Db 103 ACAGCTGAAGCAGGATGGGACCAAGTTGTGAGCTGCAGATTGTCGCTGTGCATAGCAG 162
  QY 63 ATGCTGGGAGTACTCTGTGTGTGTTGGGAGGAGGACCTCTGCCACTCTCCACGTGA 122
  Db 163 ATGCTGGAGATCTGTCGTCATATGTGAGCAGGAGGAGACCTCGGCCACGCTCACTGTCT 222
  QY 123 AGCCCTGCCACCAAGTTACAGAGGCTCTGAGAGGATGAAGAGCCCTGTGAAGGGGCCA 182
  Db 223 GGGCCCTCCGCGCCAGATTCATAGAATATGTGAGAAATCAGAGGCCACAGAGGGGCCA 282
  QY 183 CAGCCATGTTGTGTGTAACCTGAGCAAGGTGGCCCTGTGAGTGGAGGAGGGGCCCG 242
  Db 283 CAGCTGTCTGCAGTGTGAGCTGAGCAGGCGGCCCTGCGTGGATGGCGGAGGGGTCTG 342
  QY 243 AGAACCTCAGAGATGGGAC 262
  Db 343 AGACCTCAGAGCTGAGGAC 362

RESULT 10
BF826489/c
LOCUS
  PMO-OT0232-240501-005-a12 OT0232 Homo sapiens cDNA, mRNA sequence.
  EST 20-MAY-2002
  BF826489
  BF826489.1 GI:21031179
  EST.
  KEYWORDS
  SOURCE
  ORGANISM
    Homo sapiens (human)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 324)
  Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
  Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
  Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
  Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
  O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
  Simpson,A.J.J.
  Shotgun sequencing of the human transcriptome with ORF expressed
  sequence tags
  Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  20202663
  10737800
  Contact: Simpson A.J.G.
  Laboratory of Cancer Genetics
  Ludwig Institute for Cancer Research
  Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
  Brazil
  Tel: +55-11-2704922
  Fax: +55-11-2707001
  Email: asimpson@ludwig.org.br
  This sequence was derived from the FAPESP/LICR Human Cancer Genome
  Project. This entry can be seen in the following URL
  (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM4&t2=CM4-HN0020-
  221100-452-b09&t3=2000-11-22&t4=1)
  Seq primer: puc 18 forward

```

Email: asimpson@ludwig.org.br  
 This sequence was derived from the PAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PMO&t2=PMO-070232-240501-005-a12&t3=2001-05-24&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 21  
 High quality sequence stop: 324.

#### FEATURES

Location/Qualifiers  
 1..324  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="GT0232"  
 /note="Organ: ovary; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

#### ORIGIN

Query Match 43.0%; Score 165.4; DB 13; Length 324;  
 Best Local Similarity 87.4%; Pred. No. 5.1e-26;  
 Matches 181; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
 179 GCCACAGCAGTCTTGTGTGTAAGTGAAGTGGCCCTCTGCAAGTGGAGAGGGG 238  
 324 GCCACGGCCAGCTCGGTGTGAGTGAAGCAGAGCCCTGTGGAGTGGAGAAAGGG 265  
 239 CCCGAGACCTCAGAGATGGGACAGATACATCTCTGAGGACAGAGGGACCAAGTGTGAG 298  
 264 TCCGAGACCTCAGAGATGGGACAGATCTCTGAGGACAGAGGGGCCCATGTGTGAG 205  
 299 CTGCAAGATCTGTGGCTGCGCCATGGCGACGCGCGGGAGTACTTGTGTGTGGGGCAG 358  
 204 CTGCAAGATCCGTGGCTGCGCCATGGTGGATGCGCGGAGTACTCTGTGTGTGGAGAG 145  
 359 GAGAGACCTCAGCAGCGCTCACCATC 385  
 144 GAGAGACCTCAGCAGCGCTCACCATC 118

#### RESULT 11

BF398067/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 BF398067 601 bp mRNA linear EST 27-NOV-2000  
 UI-R-BS2-beg-h-06-0-UI.s1 UI-R-BS2 Rattus norvegicus cDNA clone  
 UI-R-BS2-beg-h-06-0-UI.3', mRNA sequence.  
 BF398067  
 EST.  
 Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 601)  
 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 97044477  
 8889548  
 Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4155 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. The sequence tag present in the cDNA between the NotI site  
 and the oligo-dT track served to verify it as a clone from the  
 normalized embryo at 13 dpc library cDNA Library Preparation; M.B.  
 Soares Lab Clone distribution: clones will be available through  
 Research Genetics (www.resgen.com)  
 Seq primer: M13 Forward  
 POLYA=Yes.

#### FEATURES

Location/Qualifiers  
 1..601  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strains="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-BS2-beg-h-06-0-UI"  
 /dev\_stage="embryonic 13 dpc"  
 /lab\_host="PH108 (Life Technologies)"  
 /clone\_lib="UI-R-BS2"  
 /note="Vector: pTV73D-Pac (pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-BS2  
 library is a subtracted library derived from 13 dpc whole  
 embryo tissue. For a detailed description of the library  
 from which this clone was derived, please visit our web  
 site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been  
 previously described in (Bonaldo, Lennon and Soares,  
 Genome Research 6:791-806, 1996)  
 TAG TISSUE=embryo at 13 dpc  
 TAG LIB=UI-R-BS2  
 TAG\_SEQ=AATCC"

#### ORIGIN

Query Match 42.4%; Score 163.2; DB 10; Length 601;  
 Best Local Similarity 69.4%; Pred. No. 2.1e-25;  
 Matches 222; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
 66 CTGGGAGTACTCTGTGTGTGGGAGGAGAGGACCTCTGCCACTCTCACCTGAAGG 125  
 517 CTAGGGCGTCCATGTCTCAGAGGGTGTGATTGTGTCGTGCTTCTTCTGACCTCC 458  
 126 CCCTGGCCAGCAAGTTCACAGAGGGTCTGAGGAATGAGAGCGCGTGGAGGGCCACAG 185  
 457 CCCTTCGGCTTAAGTTCATAGAAAGTCTGAGGAATGAGGGGCCACAGAGGAACACAG 398  
 186 CCATGTGTGTGTGTAAGTCTGAGCAAGTGGCCCTCTGGAGTGGAGAGGGGCCGAGA 245  
 397 CCAGCTGAGCTGCAAACTGAGCAAGCGCTCCGTGACGTGGAAGAGGAGAACAGA 338  
 246 ACCTCAGAGATGGGACAGATACATCTGAGGAGAGGGGACAGAGTGTGAGCTGAGA 305  
 337 CCTTGGAGACGGAGACAAATATGGCTGAGGACGAGGAGCTGTGTGTGAGCTGAGA 278  
 306 TCTGTGGCTGGCCATGGCGAGCCCGGGAGTACTTGTGTGTGGGGCAGAGAGGA 365  
 277 TCCGTGGCTTACACAGCCGATCTGGGGAGTACTCATGTGTGTGGCAGAGAGA 218  
 366 CCTCAGCCAGCTCACCATC 385  
 217 CATCAGTGTCTGACTGTC 198

#### RESULT 12

BH025901/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 BH025901 690 bp DNA linear GSS 17-JUL-2001  
 RPCI-24-259H14.TV RPCI-24 Mus musculus genomic clone  
 RPCI-24-259H14, genomic survey sequence.  
 BH025901  
 GSS.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE
AUTHORS      Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
              Tsugeye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
              Russell,D., de Jong,P. and Fraser,C.M.
TITLE        Mouse BAC End Sequences from Library RPCI-24
JOURNAL      Unpublished (1999)
COMMENT      Other GSSs: RPCI-24-259H14.TJ
              Contact: Shaying Zhao
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Drive, Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: szhao@tigr.org
              Clones are derived from the mouse BAC library RPCI-24. For BAC
              library availability, please contact Pieter de Jong
              (pdejong@mail.cho.org). Clones may be purchased from BACPAC
              Resources (http://www.choi.org/bacpac/orderingframe.html). BAC end
              page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
              Plate: 259 row: H column: 14
              Seq primer: T7
              Class: BAC ends.

FEATURES
source      Location/Qualifiers
            1..690
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="RPCI-24-259H14"
            /sex="Male"
            /cell_type="Spleen/Brain"
            /clone_lib="RPCI-24"
            /note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI.
            RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
            library was cloned in the pTARBAC1 cloning vector at the
            BamHI sites using MboI partially digested male C57BL/6J
            DNA."

ORIGIN
Query Match      41.2%; Score 158.8; DB 28; Length 690;
Best Local Similarity 79.1%; Pred. No. 2e-24;
Matches 201; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 125 GCCCTGCCAGCCAGTTCACAGAGGCTGAGGATGAGAGGCGCTGGAAGGGCCAC 184
DB 253 GCCCTGCCCTCCAGATTCATGAGAGATTGAGAGGCTCAAGAGGCCACAGAGGCCAATG 194

QY 185 GCCATTGTGTGTGTAAGTGAAGTGGCCCTGTGGAGTGGAGGAGGAGGCGCCGAG 244
DB 193 GCCACTCTGAGTGGCCAGATGAGCAGACTGCCCTCTGTGAGTGGAGAGGGTCTGAG 134

QY 245 AACCTCAGAGTGGGACAGATACATCTGAGGAGGAGGAGGACCAAGTGTGAGCTGCAG 304
DB 133 ACCCTGAGAGATGGGGCAGATACAGCCTGAGGAGGATGGGCTGTGTGAGCTGCAG 74

QY 305 ATCTGTGCGCTGGCCATGTCGCGGAGCGCGGAGTACTTGTGTGTCGGGAGGAGAGG 364
DB 73 ATCTGTGACCTGGTGTGCGAGAGAGCTGGGAGTACTCATGTGTGTG-GGGCAGGAGAG 15

QY 365 ACCTCAGCCAGCT 378
DB 14 ACCTCAGCCACACT 1

RESULT 13
LOCUS      CE078765      505 bp      DNA      linear      GSS 24-SEP-2003
DEFINITION tigr-gss-dog-17000324480165 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION  CE078765
VERSION     CE078765.1 GI:35145465
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)

ORGANISM    Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1 (bases 1 to 505)
AUTHORS    Kirkness,E.F., Rafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
TITLE      The dog genome: survey sequencing and comparative analysis
JOURNAL    Science 301 (5641), 1898-1903 (2003)
MEDLINE    22875432
PUBMED     14512627
COMMENT    Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun
            Location/Qualifiers
            1..505
            /organism="Canis familiaris"
            /mol_type="genomic DNA"
            /strain="Standard Poodle"
            /db_xref="taxon:9615"
            /clone_lib="Dog Library"
            /note="Site 1: BstXI; Libraries were prepared from
            peripheral blood"

ORIGIN
Query Match      33.0%; Score 127.2; DB 29; Length 505;
Best Local Similarity 83.7%; Pred. No. 1.4e-17;
Matches 144; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 214 GCGCCCTGTGGAGTGGAGAGGGCCGAGAACCTCAGAGATGGGACAGATACCT 273
DB 2 GCGCCCTGTGGAGTGGAGAGGGCCCTCAGAGCCCTCAGAGCGGGACAGGTCAGCCT 61

QY 274 GAGGACAGAGGGGACAGGTGTGAGCTGACATCTGCAGATCTGCCTGGCCATGGCGGCGCGG 333
DB 62 GAGGACAGAGGGGCGGTGTGCGAGCTGCGAGATCCGGGCTTGGCCATGGAGGACTCCGG 121

QY 334 GGAGTACTTGTGTGTGTGCGGCGAGGAGGAGGACCTCAGCCACGCTCACCATC 385
DB 122 GGAGTACTCGTCCGTGTGTGGGAGGAGGAGGCGCGCTGCTCACCGTC 173

RESULT 14
LOCUS      BG987697/c      204 bp      mRNA      linear      EST 13-JUN-2001
DEFINITION PM3-HT1165-110101-001-e07 HT1165 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BG987697
VERSION     BG987697.1 GI:14391767
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 204)
AUTHORS    Dias Neto,E., Garcia Correa,R., Vertovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsuoka,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., Geoliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
PUBMED     10737800
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research

```

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtm2.pl?t1=PM3&t2=PM3-HT1165-110101-001-e07&t3=2001-01-11&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence start: 35  
High quality sequence stop: 204.

FEATURES  
source

```

1. .204
/locus="H1165"
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="H1165"
/site2="Ogan: head neck; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from QPESNES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

## ORIGIN

	Query Match	31.8%;	Score 122.4;	DB 12;	Length 204;
	Best Local Similarity	75.0%;	Pred. No. le-16;		
	Matches 153;	Conservative 0;	Mismatches 51;	Indels 0;	Gaps 0;
QY	7	CCTGAGCAGACGCGGGCTGTGTCGAGCTGCAGATCTGTGGCCCTGGCTGTGGCAGATGC	66		
Db	204	CCTGAGCAGGACGGGACGAGTCTGAGCTGCAGATTCGTGGCCCTGGCTGTGGAGGACAC	145		
QY	67	TGGGGAGTACTCTGTGTGTGTGGGAGGAGAGAACCTCTGCCACTCTCACCGTGAAGGC	126		
Db	144	TGGAGAGTATTGTGTGTGTGTCGGCGAGGAGAGAACTCAGCTACACTCTTTCAGGGC	85		
QY	127	CCTCCAGCCAAAGTTTCAGAGGGCTCTGAGGAATGAGAGAGCCCGTGGAAAGGGGCCACAGC	186		
Db	84	CCTGCCAGCCAGATTATACCCACATGATGCCAACCGAGAGCCAGAGAGGGGCCACGGC	25		
QY	187	CATGTTTGGTGTGAACGTAGCAA	210		
Db	24	CACACTGCACGTGTGAACGTAGCAA	1		

RESULT	15
AZ590254/c	
LOCUS	
AZ590254	647 bp DNA linear GSS 13-DEC-2000
DEFINITION	Mus musculus genomic clone UGCLM039818 Mouse 10kb plasmid UGUCIM library Mus musculus genomic clone UGCLM039818 R. genomic survey sequence.

ACCESSION	AZ590254	GI:11712444
VERSION	AZ590254.1	

KEYWORDS	GSS.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus

REFERENCE 1 (bases 1 to 647)  
MUS musculus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 647)

DUNN, D.: Aovagi. A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H.: Longacre, S.: Mahmoud, M.: Meenen, E.: Pedersen, T.,

Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von

Niederhausern, A. and Wright, D.: Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

1. U.S. DEPARTMENT OF AGRICULTURE  
 2. WASHINGTON, D. C. 20250  
 3. OFFICE OF THE SECRETARY  
 4. ATTENTION: ASSISTANT SECRETARY FOR  
 5. REGISTRATION AND RECORDS  
 6. DATE: 10/10/70  
 7. TO: DIRECTOR, FBI  
 8. FROM: SECRETARY, U.S. DEPARTMENT OF AGRICULTURE  
 9. SUBJECT: RE: [REDACTED]  
 10. RE: [REDACTED]  
 11. RE: [REDACTED]  
 12. RE: [REDACTED]  
 13. RE: [REDACTED]  
 14. RE: [REDACTED]  
 15. RE: [REDACTED]  
 16. RE: [REDACTED]  
 17. RE: [REDACTED]  
 18. RE: [REDACTED]  
 19. RE: [REDACTED]  
 20. RE: [REDACTED]  
 21. RE: [REDACTED]  
 22. RE: [REDACTED]  
 23. RE: [REDACTED]  
 24. RE: [REDACTED]  
 25. RE: [REDACTED]  
 26. RE: [REDACTED]  
 27. RE: [REDACTED]  
 28. RE: [REDACTED]  
 29. RE: [REDACTED]  
 30. RE: [REDACTED]  
 31. RE: [REDACTED]  
 32. RE: [REDACTED]  
 33. RE: [REDACTED]  
 34. RE: [REDACTED]  
 35. RE: [REDACTED]  
 36. RE: [REDACTED]  
 37. RE: [REDACTED]  
 38. RE: [REDACTED]  
 39. RE: [REDACTED]  
 40. RE: [REDACTED]  
 41. RE: [REDACTED]  
 42. RE: [REDACTED]  
 43. RE: [REDACTED]  
 44. RE: [REDACTED]  
 45. RE: [REDACTED]  
 46. RE: [REDACTED]  
 47. RE: [REDACTED]  
 48. RE: [REDACTED]  
 49. RE: [REDACTED]  
 50. RE: [REDACTED]  
 51. RE: [REDACTED]  
 52. RE: [REDACTED]  
 53. RE: [REDACTED]  
 54. RE: [REDACTED]  
 55. RE: [REDACTED]  
 56. RE: [REDACTED]  
 57. RE: [REDACTED]  
 58. RE: [REDACTED]  
 59. RE: [REDACTED]  
 60. RE: [REDACTED]  
 61. RE: [REDACTED]  
 62. RE: [REDACTED]  
 63. RE: [REDACTED]  
 64. RE: [REDACTED]  
 65. RE: [REDACTED]  
 66. RE: [REDACTED]  
 67. RE: [REDACTED]  
 68. RE: [REDACTED]  
 69. RE: [REDACTED]  
 70. RE: [REDACTED]  
 71. RE: [REDACTED]  
 72. RE: [REDACTED]  
 73. RE: [REDACTED]  
 74. RE: [REDACTED]  
 75. RE: [REDACTED]  
 76. RE: [REDACTED]  
 77. RE: [REDACTED]  
 78. RE: [REDACTED]  
 79. RE: [REDACTED]  
 80. RE: [REDACTED]  
 81. RE: [REDACTED]  
 82. RE: [REDACTED]  
 83. RE: [REDACTED]  
 84. RE: [REDACTED]  
 85. RE: [REDACTED]  
 86. RE: [REDACTED]  
 87. RE: [REDACTED]  
 88. RE: [REDACTED]  
 89. RE: [REDACTED]  
 90. RE: [REDACTED]  
 91. RE: [REDACTED]  
 92. RE: [REDACTED]  
 93. RE: [REDACTED]  
 94. RE: [REDACTED]  
 95. RE: [REDACTED]  
 96. RE: [REDACTED]  
 97. RE: [REDACTED]  
 98. RE: [REDACTED]  
 99. RE: [REDACTED]  
 100. RE: [REDACTED]

REVISED  
PUBLISHED (2000)

JOURNAL  
 Unpublished (2006)  
 Contact: Robert B Weiss  
 COMMENT

CONTACT: ROBERT D. HEID  
University of Utah Genome Center

UNIVERSITY OF UTAH  
UNIVERSITY OF UTAH

UNIVERSITY OF OREGON  
PM 308 Biomedical Polymers Research Bldg., 20 S. 2030 E., SL

ALL INFORMATION CONTAINED HEREIN IS UNCLASSIFIED

84112 USA  
Tel: 801 585 5506  
Fax: 801 585 7177  
Email: [cdunne@genetics.utah.edu](mailto:cdunne@genetics.utah.edu)  
Insert Length: 10000 Std Error: 0.00  
Plate: 0399 row: B column: 18  
Seq primer: CACAGGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 647.

**FEATURES**

1. .647

```
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0399B18"
/sex="Male"
/lab_host="F. Coli strain XL10-Gold, Tl-resistant, F-"
/clone_lib="Mouse 10kb UUGC1M library"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydronically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 [G14732114|gb|AF129072.1], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed into adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
```

ORIGIN:

	Query Match	31.6%	Score 121.8	DB 28	Length 647
	Best Local Similarity	79.6%	Pred. No. 2.3e-16		
	Matches 144	Conservative 0	Mismatches 37	Indels 0	Gaps 0
QY	125	GCCTCGCCAGCCAAATTACAGAGGGTCTGAGGAATGAAGAGGCCGTGGAGAGGGGCCACA	184		
Db	181	GGCTTGCTGCCAAATTATAGAAGATCTGAGAAGCCCAAGAAGCCACAGAAGAGGGGCCACA	122		
QY	185	GCCATGTTGTGTTGATGAACAGAGGTGGCCCTGTGGNATCGAGGAAGGGCCCGAG	244		
Db	121	GCAATCTTAAGATGTGAGCTGAGCAAGGCTGCTCCTGTGGAGTGGAGGAAGGGTCTGAG	62		
QY	245	AACCTCAGAGATGGGACAGATACATCTCGAGGACAGGAGGGACCCAGGTGTGAGCTGCAG	304		
Db	61	ACCTTGAAGATGGGACAGATACACCTCGAGGACGATGGGCCGTATGTGAGCTGCAG	2		
QY	305	A 305			
Db	1	A 1			

Search completed: September 19, 2004, 22:07:30  
Job time : 1127.53 secs

**This Page Blank (uspto)**

Result No.	Score	Query		ID	Description
		Match	Length		
1	385	100.0	7564	17	US-10-476-397-15
2	385	100.0	14061	16	US-10-093-463-73
3	385	100.0	14109	15	US-10-093-463-71
4	385	100.0	23907	14	US-10-077-130-6
5	385	100.0	24120	14	US-10-077-130-4
6	70	18.2	2534	16	US-10-108-260A-1946
7	46.8	12.2	81940	9	US-09-759-508B-1
8	46.8	12.2	81940	10	US-09-960-706-1092
9	46.8	12.2	81940	10	US-09-873-319-724
10	46.2	12.0	60	10	US-09-908-975-13516
11	43.4	11.3	671	15	US-10-184-644-346
12	43.4	11.3	671	15	US-10-184-634-346
13	43.4	11.3	1645	9	US-09-726-643-22
14	43.4	11.3	1645	14	US-10-042-141-22
					Sequence 15, Appl
					Sequence 73, Appl
					Sequence 71, Appl
					Sequence 6, Appl
					Sequence 4, Appl
					Sequence 1946, Ap
					Sequence 1, Appl
					Sequence 1092, Ap
					Sequence 724, Appl
					Sequence 13516, A
					Sequence 346, App
					Sequence 346, App
					Sequence 22, Appl
					Sequence 22, Appl

;; PRIOR FILING DATE: 2002-05-01  
;; PRIOR APPLICATION NUMBER: US 60/288,290  
;; PRIOR FILING DATE: 2001-05-02  
;; PRIOR APPLICATION NUMBER: US 60/292,468  
;; PRIOR FILING DATE: 2001-05-21  
;; PRIOR APPLICATION NUMBER: US 60/298,616  
;; PRIOR FILING DATE: 2001-06-15  
;; PRIOR APPLICATION NUMBER: US 60/301,672  
;; PRIOR FILING DATE: 2001-06-28  
;; PRIOR APPLICATION NUMBER: US 60/345,008  
;; PRIOR FILING DATE: 2002-01-04  
;; NUMBER OF SEQ ID NOS: 22  
;; SOFTWARE: PERL Program  
;; SEQ ID NO 15  
;; LENGTH: 7564  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; OTHER INFORMATION: Incyte ID No: 7326129CB1  
US-10-476-397-15

Query Match 100.0%; Score 385; DB 17; Length 7564;  
Best Local Similarity 100.0%; Pred. No. 7e-98;  
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACATAGCCTCAGCAGACGGGCTGTGTCGAGCTGCAGATCTGTGGCTGTGGCTGTGGC 60  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
4332 ACATAGCCTCAGCAGACGGGCTGTGTCGAGCTGCAGATCTGTGGCTGTGGCTGTGGC 4391

Qy 61 AGATGCTGGGAGTACTCTCTGTGTGTGGGAGGAGACCTCTGCCACTCTCACCGT 120  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
4392 AGATGCTGGGAGTACTCTCTGTGTGTGGGAGGAGACCTCTGCCACTCTCACCGT 4451

Qy 121 GAAGCCCTCCAGCCAGTTCACAGAGGCTGTGAGGATGAAGAGCGCTGGAGGGGC 180  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
4452 GAAGCCCTCCAGCCAGTTCACAGAGGCTGTGAGGATGAAGAGCGCTGGAGGGGC 4511

Qy 181 CACAGCATGTGTGGTGTGAATCTGAGCAAGTGGCCCTGTGGAGTGGAGAGGGGCC 240  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
4512 CACAGCATGTGTGGTGTGAATCTGAGCAAGTGGCCCTGTGGAGTGGAGAGGGGCC 4571

Qy 241 CGAGAACTCAGAGATGGGACAGATACATCTTGAAGGAGGAGGACACAGGTGTGAGCT 300  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
4572 CGAGAACTCAGAGATGGGACAGATACATCTTGAAGGAGGAGGACACAGGTGTGAGCT 4631

Qy 301 GCAGATCTGTGGCTGSCCATGCGGACCGCGGGGAGTACTTGTGTGTGGGGCAGGA 360  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
4632 GCAGATCTGTGGCTGSCCATGCGGACCGCGGGGAGTACTTGTGTGTGGGGCAGGA 4691

Qy 361 GAGGACCTCAGCAGCTCACCATC 385  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
4692 GAGGACCTCAGCAGCTCACCATC 4716

RESULT 2  
US-10-093-463-73  
;; Sequence 73, Application US/10093463  
;; Publication No. US20030208039A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Padigar, Muralidhara  
;; APPLICANT: Shenoy, Suresh  
;; APPLICANT: Kekuda, Ramesh  
;; APPLICANT: Gusev, Vladimir  
;; APPLICANT: Pochart, Pascal  
;; APPLICANT: Zhong, Mei  
;; APPLICANT: Rastelli, Luca  
;; APPLICANT: Mezes, Peter  
;; APPLICANT: Smithson, Glenda  
;; APPLICANT: Guo, Xiaojia  
;; APPLICANT: Gerlach, Valerie  
;; APPLICANT: Casman, Stacie  
;; APPLICANT: Boldog, Ferenc

;; APPLICANT: Li, Li  
;; APPLICANT: Zerhusen, Bryan  
;; APPLICANT: Tchernev, Velizar  
;; APPLICANT: Gangolli, Esha  
;; APPLICANT: Vernet, Corine  
;; APPLICANT: Pena, Carol  
;; APPLICANT: Burgess, Catherine  
;; APPLICANT: Liu, Xiaohong  
;; APPLICANT: Spytek, Kimberly  
;; APPLICANT: Gorman, Linda  
;; APPLICANT: Spaderna, Steven  
;; APPLICANT: Voss, Edward  
;; APPLICANT: Malyankar, Uriel  
;; APPLICANT: Anderson, David  
;; APPLICANT: Patturajan, Meera  
;; APPLICANT: Miller, Charles  
;; APPLICANT: Taupier, Raymond J. Jr.  
;; TITLE OF INVENTION: No. US20030208039A1 Antibodies that Bind to Antigenic Polypeptides  
;; FILE REFERENCE: 21402-290A (Cura 590AT)  
;; CURRENT APPLICATION NUMBER: US/10/093,463  
;; CURRENT FILING DATE: 2002-06-24  
;; PRIOR APPLICATION NUMBER: 60/283,675  
;; PRIOR FILING DATE: 2001-04-14  
;; PRIOR APPLICATION NUMBER: 60/338,092  
;; PRIOR FILING DATE: 2001-12-03  
;; PRIOR APPLICATION NUMBER: 60/274,281  
;; PRIOR FILING DATE: 2001-03-08  
;; PRIOR APPLICATION NUMBER: 60/274,101  
;; PRIOR FILING DATE: 2001-03-08  
;; PRIOR APPLICATION NUMBER: 60/325,681  
;; PRIOR FILING DATE: 2001-09-27  
;; PRIOR APPLICATION NUMBER: 60/304,354  
;; PRIOR FILING DATE: 2001-07-10  
;; PRIOR APPLICATION NUMBER: 60/279,995  
;; PRIOR FILING DATE: 2001-03-30  
;; PRIOR APPLICATION NUMBER: 60/294,999  
;; PRIOR FILING DATE: 2001-05-31  
;; PRIOR APPLICATION NUMBER: 60/287,424  
;; PRIOR FILING DATE: 2001-04-30  
;; PRIOR APPLICATION NUMBER: 60/299,027  
;; PRIOR FILING DATE: 2001-06-18  
;; PRIOR APPLICATION NUMBER: 60/309,198  
;; PRIOR FILING DATE: 2001-07-31  
;; PRIOR APPLICATION NUMBER: 60/281,194  
;; PRIOR FILING DATE: 2001-04-04  
;; PRIOR APPLICATION NUMBER: 60/274,194  
;; PRIOR FILING DATE: 2001-03-08  
;; PRIOR APPLICATION NUMBER: 60/274,849  
;; PRIOR FILING DATE: 2001-03-09  
;; PRIOR APPLICATION NUMBER: 60/330,380  
;; PRIOR FILING DATE: 2001-10-18  
;; PRIOR APPLICATION NUMBER: 60/275,235  
;; PRIOR FILING DATE: 2001-03-12  
;; PRIOR APPLICATION NUMBER: 60/288,342  
;; PRIOR FILING DATE: 2001-05-03  
;; PRIOR APPLICATION NUMBER: 60/275,578  
;; PRIOR FILING DATE: 2001-03-13  
;; NUMBER OF SEQ ID NOS: 370  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 73  
;; LENGTH: 14061  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (15) .. (14039)  
US-10-093-463-73

Query Match 100.0%; Score 385; DB 16; Length 14061;  
Best Local Similarity 100.0%; Pred. No. 7e-98;  
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Qy 1 ACATAGCTGAGGACGAGCGGCTGTGTGCGAGCTGCAGATCTGTGGCCCTGTGGC 60  
Db 2447 ACATAGCTGAGGACGAGCGGCTGTGTGCGAGCTGCAGATCTGTGGCCCTGTGGC 2506  
Qy 61 AGATGCTGGGAGTACTCTGTGTGTGTGGGAGGAGGACCTCTGCCACTCTCACCGT 120  
Db 2507 AGATGCTGGGAGTACTCTGTGTGTGTGGGAGGAGGACCTCTGCCACTCTCACCGT 2566  
Qy 121 GAAGGCCCTGCACCCAGTTTACAGAGGGCTCTGAGGAATGAAGAGCCCGTGAAGGGGC 180  
Db 2567 GAAGGCCCTGCACCCAGTTTACAGAGGGCTCTGAGGAATGAAGAGCCCGTGAAGGGGC 2626  
Qy 181 CACAGCCATGTTGTGTGTGAACTGACAAAGGTGGCCCTGTGGAGTGGAGGAGGGGCC 240  
Db 2627 CACAGCCATGTTGTGTGTGAACTGACAAAGGTGGCCCTGTGGAGTGGAGGAGGGGCC 2686  
Qy 241 CGAGAACCTCAGAGATGGGACAGATACATCTTGAGGACGAGGAGGACGAGTGTGAGCT 300  
Db 2687 CGAGAACCTCAGAGATGGGACAGATACATCTTGAGGACGAGGAGGACGAGTGTGAGCT 2746  
Qy 301 GCAGATCTGTGGCTGCGCATGGCGGACGCGGGAGTACTTGTGTGTGTGGCGGACGGA 360  
Db 2747 GCAGATCTGTGGCTGCGCATGGCGGACGCGGGAGTACTTGTGTGTGTGGCGGACGGA 2806  
Qy 361 GAGGACCTCAGCCACGCTCACCATC 385  
Db 2807 GAGGACCTCAGCCACGCTCACCATC 2831

### RESULT 3

US-10-093-463-71  
; Sequence 71, Application US/10093463  
; Publication No. US20030208039A1  
; GENERAL INFORMATION:  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Gusev, Vladimir  
; APPLICANT: Pochart, Pascal  
; APPLICANT: Zhong, Mei  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Mezes, Peter  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Casman, Stacie  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Li, Li  
; APPLICANT: Zerhusen, Bryan  
; APPLICANT: Tchernev, Velizar  
; APPLICANT: Gangolli, Esba  
; APPLICANT: Vernet, Corine  
; APPLICANT: Pena, Carol  
; APPLICANT: Burgess, Catherine  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Gorman, Linda  
; APPLICANT: Spaderna, Steven  
; APPLICANT: Voss, Edward  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Anderson, David  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Miller, Charles  
; APPLICANT: Taupier, Raymond J. Jr.  
; TITLE OF INVENTION: No. US20030208039A1el Antibodies that Bind to Antigenic Polypepti  
; FILE REFERENCE: 21402-290A (Cura 550A)  
; CURRENT APPLICATION NUMBER: US/10/093,463  
; CURRENT FILING DATE: 2002-06-24  
; PRIOR APPLICATION NUMBER: 60/283,675  
; PRIOR FILING DATE: 2001-04-14  
; PRIOR APPLICATION NUMBER: 60/338,092  
; PRIOR FILING DATE: 2001-12-03

; PRIOR APPLICATION NUMBER: 60/274,281  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/274,101  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/325,681  
; PRIOR FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: 60/304,354  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/279,995  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 60/294,899  
; PRIOR FILING DATE: 2001-05-31  
; PRIOR APPLICATION NUMBER: 60/287,424  
; PRIOR FILING DATE: 2001-04-30  
; PRIOR APPLICATION NUMBER: 60/299,027  
; PRIOR FILING DATE: 2001-06-18  
; PRIOR APPLICATION NUMBER: 60/309,198  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/281,194  
; PRIOR FILING DATE: 2001-04-04  
; PRIOR APPLICATION NUMBER: 60/274,194  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/274,849  
; PRIOR FILING DATE: 2001-03-09  
; PRIOR APPLICATION NUMBER: 60/330,380  
; PRIOR FILING DATE: 2001-10-18  
; PRIOR APPLICATION NUMBER: 60/275,235  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: 60/288,342  
; PRIOR FILING DATE: 2001-05-03  
; PRIOR APPLICATION NUMBER: 60/275,578  
; PRIOR FILING DATE: 2001-03-13  
; NUMBER OF SEQ ID NOS: 370  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 71  
; LENGTH: 14109  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (15)..(14088)  
US-10-093-463-71  
Query Match 100.0%; Score 385; DB 16; Length 14109;  
Best Local Similarity 100.0%; Pred. No. 7e-98;  
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ACATAGCTGAGGACGAGCGGCTGTGTGCGAGCTGCAGATCTGTGGCCCTGTGGC 60  
Db 2447 ACATAGCTGAGGACGAGCGGCTGTGTGCGAGCTGCAGATCTGTGGCCCTGTGGC 2506  
Qy 61 AGATGCTGGGAGTACTCTGTGTGTGTGGGAGGAGGACCTCTGCCACTCTCACCGT 120  
Db 2507 AGATGCTGGGAGTACTCTGTGTGTGTGGGAGGAGGACCTCTGCCACTCTCACCGT 2566  
Qy 121 GAAGGCCCTGCACCCAGTTTACAGAGGGTCTGAGGAATGAAGAGCCCGTGAAGGGGC 180  
Db 2567 GAAGGCCCTGCACCCAGTTTACAGAGGGTCTGAGGAATGAAGAGCCCGTGAAGGGGC 2626  
Qy 181 CACAGCCATGTTGTGTGTGAACTGACAAAGGTGGCCCTGTGGAGTGGAGGAGGGGCC 240  
Db 2627 CACAGCCATGTTGTGTGTGAACTGACAAAGGTGGCCCTGTGGAGTGGAGGAGGGGCC 2686  
Qy 241 CGAGAACCTCAGAGATGGGACAGATACATCTTGAGGACGAGGAGGACGAGTGTGAGCT 300  
Db 2687 CGAGAACCTCAGAGATGGGACAGATACATCTTGAGGACGAGGAGGACGAGTGTGAGCT 2746  
Qy 301 GCAGATCTGTGGCTGCGCATGGCGGACGCGGGAGTACTTGTGTGTGTGGCGGACGGA 360  
Db 2747 GCAGATCTGTGGCTGCGCATGGCGGACGCGGGAGTACTTGTGTGTGTGGCGGACGGA 2806  
Qy 361 GAGGACCTCAGCCACGCTCACCATC 385

```
Db 2807 GAGGACCTCAGCCACGCTCACCATC 2831

RESULT 4
US-10-077-130-6
; Sequence 6, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047P1RCP1(N)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-6

Query Match 100.0%; Score 385; DB 14; Length 23907;
Best Local Similarity 100.0%; Pred. No. 7e-98;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACATAGCTGAGGACGAGCGGGCTGTGCGAGCTGCAGATCTGTGGCTGGCTGTGGC 60
Db 10215 ACATAGCTGAGGACGAGCGGGCTGTGCGAGCTGCAGATCTGTGGCTGGCTGTGGC 10274
QY 61 AGATGCTGGGAGPACTCTCTGTGTGTGTGGGAGGAGACCTCTGCCACTCTCACCGT 120
Db 10275 AGATGCTGGGAGPACTCTCTGTGTGTGTGGGAGGAGACCTCTGCCACTCTCACCGT 10334
QY 121 GAGGCGCTCCAGCCCAAGTTCACAGAGGCTCTGAGGATGAGAGCGCGTGAAGGGC 180
Db 10335 GAGGCGCTCCAGCCCAAGTTCACAGAGGCTCTGAGGATGAGAGCGCGTGAAGGGC 10394
QY 181 CACAGCCATGTTGTGTGTGAACCTGAGCAAGGTGGCCCTGTGGAGTGGAGGAAGGGCC 240
Db 10395 CACAGCCATGTTGTGTGTGAACCTGAGCAAGGTGGCCCTGTGGAGTGGAGGAAGGGCC 10454
QY 241 CGAGAACCTCAGAGTGGGACAGATACATCTTGGAGGAGGAGGACGAGTGTGAGCT 300
Db 10455 CGAGAACCTCAGAGTGGGACAGATACATCTTGGAGGAGGAGGACGAGTGTGAGCT 10514
QY 301 GCAGATCTGTGGCTGGCCATGGCGACGCGGGGAGTACTTGTGTGTGTGGCGGAGGA 360
Db 10515 GCAGATCTGTGGCTGGCCATGGCGACGCGGGGAGTACTTGTGTGTGTGGCGGAGGA 10574
QY 361 GAGGACCTCAGCCACGCTCACCATC 385
Db 10575 GAGGACCTCAGCCACGCTCACCATC 10599

RESULT 5
US-10-077-130-4
; Sequence 4, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047P1RCP1(N)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 24120
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-4

Query Match 100.0%; Score 385; DB 14; Length 24120;
Best Local Similarity 100.0%; Pred. No. 7e-98;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACATAGCTGAGGACGAGCGGGCTGTGCGAGCTGCAGATCTGTGGCTGGCTGTGGC 60
Db 10286 ACATAGCTGAGGACGAGCGGGCTGTGCGAGCTGCAGATCTGTGGCTGGCTGTGGC 10345
QY 61 AGATGCTGGGAGTACTCTCTGTGTGTGTGGGAGGAGGACCTCTGCCACTCTCACCGT 120
Db 10346 AGATGCTGGGAGTACTCTCTGTGTGTGTGGGAGGAGGACCTCTGCCACTCTCACCGT 10405
QY 121 GAAGCCCTGCCAGCCAAAGTTCACAGAGGCTCTGAGGAATGAAGAGCGCGTGAAGGGC 180
Db 10406 GAAGCCCTGCCAGCCAAAGTTCACAGAGGCTCTGAGGAATGAAGAGCGCGTGAAGGGC 10465
QY 181 CACAGCCATGTTGTGTGTGAACCTGAGCAAGGTGGCCCTGTGGAGTGGAGGAAGGGCC 240
Db 10466 CACAGCCATGTTGTGTGTGAACCTGAGCAAGGTGGCCCTGTGGAGTGGAGGAAGGGCC 10525
QY 241 CGAGAACCTCAGAGTGGGACAGATACATCTTGGAGGAGGAGGACGAGTGTGAGCT 300
Db 10526 CGAGAACCTCAGAGTGGGACAGATACATCTTGGAGGAGGAGGACGAGTGTGAGCT 10585
QY 301 GCAGATCTGTGGCTGGCCATGGCGACGCGGGGAGTACTTGTGTGTGTGGCGGAGGA 360
Db 10586 GCAGATCTGTGGCTGGCCATGGCGACGCGGGGAGTACTTGTGTGTGTGGCGGAGGA 10645
QY 361 GAGGACCTCAGCCACGCTCACCATC 385
Db 10646 GAGGACCTCAGCCACGCTCACCATC 10670

RESULT 6
US-10-108-260A-1946
; Sequence 1946, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1946
; LENGTH: 2534
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1946

Query Match 18.2%; Score 70; DB 16; Length 2534;
Best Local Similarity 51.4%; Pred. No. 5.7e-10;
Matches 188; Conservative 0; Mismatches 175; Indels 3; Gaps 1;

QY 17 GACGGGGCTGTGTCGAGCTGCAGATCTGTGGCTGTGGAGTGTGGAGTGTGGAGTAC 76
Db 937 GAGGGACGATGGCCATGCTGTGTATCCGGGGCCCTCGCTCAAGGACGCGGCGAGTAC 996
```

QY 77 TCCTGTGTGTGGGAGGAGGACCTCTGCGACTCTCACCCTGAAGCCCTGCCAGCC 136  
 Db 997 ACGTGTAGGTGAGGCTTCCAGAGACAGCCAGCTCTCATGTGGAAGAAAGCAAC 1056  
 QY 137 AAGTTCACAGAGGCTCTGAGGAATGAAGAGCCGCTGAAAGGGCCACAGCCATGTTGTGG 196  
 Db 1057 TGCTTCACAGAGGCTGACCAATCTGCAGGTGGAGGAAAGGCACAGCTGTGTTCAAG 1116  
 QY 197 TGTGA---ACTGAGCAAGTGGCCCTGTGAGTGGAGGAAGGGCCCGAGAACCTCAGA 253  
 Db 1117 TGAAGACGAGACCCCGGGCCACAGTACCTGGCGAAGGCTCTTGGAGCTACGG 1176  
 QY 254 GATGGGACAGATACATCTCTGAGCGAGGGGACCAAGGTGTGAGCTGCAGATCTCTGCG 313  
 Db 1177 GCCTCAGGAAGACACAGCCAGCCAGGAGGCTGACCCCTGGGTCCACCATCACTGCC 1236  
 QY 314 CTGGCCATGCGGACGCGCGGGAGTACTTGTGTGTGTCGGGACAGAGGACCTCAGCC 373  
 Db 1237 CTGGAGAGGCAGACAGCGACACCTATACCTGCGACATTGGCCAGGCCCCAGTCCCGGGCC 1296  
 QY 374 ACGCTC 379  
 Db 1297 CAGCTC 1302

RESULT 7  
 US-09-759-508B-1  
 ; Sequence 1, Application US/09759508B  
 ; Publication No. US20020182599A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fishman, Mark C.  
 ; TITLE OF INVENTION: Methods for Diagnosing and Treating Heart Disease  
 ; FILE REFERENCE: 00786/381002  
 ; CURRENT APPLICATION NUMBER: US/09/759,508B  
 ; CURRENT FILING DATE: 2001-01-12  
 ; PRIOR APPLICATION NUMBER: US 60/175,787  
 ; PRIOR FILING DATE: 2000-01-12  
 ; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 1  
 ; LENGTH: 81940  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (133)..(80910)  
 ; OTHER INFORMATION:  
 US-09-759-508B-1

Query Match 12.2%; Score 46.8; DB 9; Length 81940;  
 Best Local Similarity 48.8%; Pred. No. 0.0017;  
 Matches 157; Conservative 0; Mismatches 162; Indels 3; Gaps 1;  
 QY 66 CTGGGGAGTACTCTCTGTGTGTGGGAGGAGAGACCTCTGCCACTCTCACCCTGAAGG 125  
 Db 16091 CAGGAGAGGTTTCCTTCAGGCTGCTAATGCCAAATCTGCAGCCAACTGAAAGTGAAG 16150  
 QY 126 CCTGCCAGCCAGTTCACAGAGGCTGAGGAATGAAGAGCCGTGGAGGGGCCACAG 185  
 Db 16151 AATTGCCCTTATCTTCATCACCTCTCAGTGATGTTAAAGTCTTCGAGAAAGATGAGG 16210  
 QY 186 CCATGTTGTGTGTAAGTGAAGTGGCC---CCTGTGGAGTGGAGGAAGGGGCCCG 242  
 Db 16211 CTAAGTTTGAAGTGAAGTATCCAGGAGACCCAAACATTCCGTTGGCTAAAGGAACCC 16270  
 QY 243 AGAACCTCAGAGATGGGACAGATACCTCTGAGGAGGGGACCAAGGTGTGAGCTGC 302  
 Db 16271 AGGAAATCACAGGTGATGACAGATTGAGCTTATAAAGATGGCACATAAGCATTCATGG 16330  
 QY 303 AGATCTGTGGCTGGCCATGGCGGACGCGGGAGTACTTGTGTGTGCGGGCAGGAGA 362  
 Db 16331 TGATCAAGTCAGCTGCTTTTGAAGATGAGCAAAATACATGTTTGAAGCTGAAGATAAGC 16390

QY 363 GGACCTCAGCCACGCTCACCAT 384  
 Db 16391 ACACAAGTGGCAAACTGATCAT 16412

RESULT 8  
 US-09-960-706-1092  
 ; Sequence 1092, Application US/09960706  
 ; Publication No. US20030134280A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mungler, William E.  
 ; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperpla  
 ; FILE REFERENCE: 44921-5029-01US  
 ; CURRENT APPLICATION NUMBER: US/09/960,706  
 ; CURRENT FILING DATE: 2001-09-24  
 ; PRIOR APPLICATION NUMBER: 60/223,323  
 ; PRIOR FILING DATE: 2000-08-07  
 ; PRIOR APPLICATION NUMBER: 09/873,319  
 ; PRIOR FILING DATE: 2001-06-05  
 ; NUMBER OF SEQ ID NOS: 1124  
 ; SOFTWARE: Patent in Ver. 2.1  
 ; SEQ ID NO 1092  
 ; LENGTH: 81940  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: Genbank Accession No. US20030134280A1 X90568  
 US-09-960-706-1092

Query Match 12.2%; Score 46.8; DB 10; Length 81940;  
 Best Local Similarity 48.8%; Pred. No. 0.0017;  
 Matches 157; Conservative 0; Mismatches 162; Indels 3; Gaps 1;  
 QY 66 CTGGGGAGTACTCTCTGTGTGTGGGAGGAGAGACCTCTGCCACTCTCACCCTGAAGG 125  
 Db 16091 CAGGAGAGGTTTCCTTCAGGCTGCTAATGCCAAATCTGCAGCCAACTGAAAGTGAAG 16150  
 QY 126 CCTGCCAGCCAGTTCACAGAGGCTGAGGAATGAAGAGCCGTGGAGGGGCCACAG 185  
 Db 16151 AATTGCCCTTATCTTCATCACCTCTCAGTGATGTTAAAGTCTTCGAGAAAGATGAGG 16210  
 QY 186 CCATGTTGTGTGTAAGTGAAGTGGCC---CCTGTGGAGTGGAGGAAGGGGCCCG 242  
 Db 16211 CTAAGTTTGAAGTGAAGTATCCAGGAGCCCAAAACATTCCGTTGGCTAAAGGAACCC 16270  
 QY 243 AGAACCTCAGAGATGGGACAGATACCTCTGAGGAGGGGACCAAGGTGTGAGCTGC 302  
 Db 16271 AGGAAATCACAGGTGATGACAGATTGAGCTTATAAAGATGGCACATAAGCATTCATGG 16330  
 QY 303 AGATCTGTGGCTGGCCATGGCGGACGCGGGAGTACTTGTGTGTGCGGGCAGGAGA 362  
 Db 16331 TGATCAAGTCAGCTGCTTTTGAAGATGAGCAAAATACATGTTTGAAGCTGAAGATAAGC 16390  
 QY 363 GGACCTCAGCCACGCTCACCAT 384  
 Db 16391 ACACAAGTGGCAAACTGATCAT 16412

RESULT 9  
 US-09-873-319-724  
 ; Sequence 724, Application US/09873319A  
 ; Publication No. US20030134324A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mungler, William E.  
 ; APPLICANT: Kulkarni, Prakash  
 ; APPLICANT: Getzenberg, Robert H.  
 ; APPLICANT: Waga, Iwao  
 ; APPLICANT: Yamamoto, Jun  
 ; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic  
 ; FILE REFERENCE: 44921-5029-US  
 ; CURRENT APPLICATION NUMBER: US/09/873,319A





Qy 359 GAGGAGACCTCAGCCAGCTCACCATC 385  
Db 1310 GAGTCGGCTCCTTCACTGTACCGTC 1336

## RESULT 15

US-10-425-114-2127/c  
; Sequence 2127, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 2127  
; LENGTH: 1275  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700204307\_FLI  
US-10-425-114-2127

Query Match 11.2%; Score 43.2; DB 13; Length 1275;  
Best Local Similarity 48.0%; Pred. No. 0.017;  
Matches 123; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Qy 121 GAAGGCCCTGCCAGCCAAAGTTCACAGAGGGTCTGAGGAATGAAGAGGCCGTGGAAGGGGC 180  
Db 443 GAATGCTCACCAAGCCATGCCCATGTGATGTGAGGGAGAGGCGGTGAAGGGGAG 384  
Qy 181 CACAGCCATGTGTGGTGTGAACGTAGCAAGGTGGCCCTGTGGAGTGGAGGAAGGGGCC 240  
Db 383 CACATTGGTGTTCATACGGAGAGGCCGAGCGCTCCACCTCGGAGAGGGCGCGATGAG 324  
Qy 241 CGAGAACTCAGAGATGGGACACATACATCCTGAGCGAGGGGACACAGGTGTGAGCT 300  
Db 323 CACCCCTTGGGTGTTCGAGTGGACCTTACACAGACCGTGGCGCCGACAGCCGGGC 264  
Qy 301 GCAGATCTGTGGCTTGGCCATGGCGGACGCCGGGGAGTACTTGTGTGTGGGGCAGGA 360  
Db 263 CTCGATCTCCGGCAGAGAGCGTGGCGAGGCTCGGGCCGCCACGAGTTGTCTCGTCAGGA 204  
Qy 361 GAGGACCTCAGCCACG 376  
Db 203 GGAGCCCTCGTCGTCG 188

Search completed: September 20, 2004, 02:51:57  
Job time : 201.275 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:09:24 : Search time 34.5752 Seconds  
(without alignments)  
6179.453 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_10286\_10670

Perfect score: 385

Sequence: 1 acatagcctgaggcaggagc.....cctcagccagctcaccatc 385

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	54.2	14.1	7218	1	US-08-232-463-14
2	44.8	11.6	289	3	US-09-007-005-17
3	44.8	11.6	289	3	US-09-244-796-17
4	40.6	10.5	7055	4	US-09-976-594-941
5	39	10.1	2277	1	US-08-676-867-5
6	39	10.1	2277	1	US-08-676-974-5
7	39	10.1	2277	2	US-09-098-487-5
8	37.6	9.8	1975	4	US-09-308-345A-6
9	37.4	9.7	9531	4	US-09-163-748C-3
10	37.2	9.7	11461	3	US-08-669-161A-29
11	37.2	9.6	2478	4	US-09-687-050-5
12	37	9.6	2478	4	US-09-921-667-15
13	37	9.6	3597	4	US-09-016-434-1319
14	37	9.6	3597	4	US-09-023-655-1314
15	37	9.6	3597	4	US-09-023-655-1314
16	36	9.4	1572	4	US-09-252-991A-3270
17	36	9.4	1965	4	US-09-252-991A-3234
18	36	9.4	2325	4	US-09-252-991A-3413
19	35.6	9.2	1926	4	US-09-249-585A-4
20	35.6	9.2	1931	2	US-09-130-114-2
21	35.4	9.2	923	4	US-09-336-643A-15
22	35.4	9.2	2612	3	US-09-042-785A-3
23	35.4	9.2	2638	3	US-09-042-785A-22
24	35.4	9.2	3474	4	US-09-527-236A-1
25	35.4	9.2	3474	4	US-09-756-854-1
26	35	9.1	584	3	US-08-959-382-3
27	35	9.1	584	4	US-09-314-844F-3

28 35 9.1 33529 3 US-09-144-085-3 Sequence 3, Appli  
29 34.6 9.0 1479 1 US-08-644-271-31 Sequence 31, Appl  
30 34.6 9.0 1479 4 US-09-077-955-35 Sequence 35, Appl  
C 31 2079 4 US-09-992-481-1 Sequence 1, Appl  
32 33.8 8.8 3277 4 US-09-963-137-163 Sequence 163, App  
C 33 33.6 8.7 5103 4 US-09-252-991A-5192 Sequence 5192, Ap  
34 33.6 8.7 6876 4 US-09-252-991A-5283 Sequence 5283, Ap  
C 35 33.4 8.7 3695 3 US-09-211-704A-1 Sequence 1, Appli  
36 33.2 8.6 2277 1 US-08-676-967-2 Sequence 2, Appli  
37 33.2 8.6 2277 1 US-08-676-974-2 Sequence 2, Appli  
38 33.2 8.6 2277 2 US-09-428-487-2 Sequence 2, Appli  
39 33.2 8.6 50937 3 US-09-428-517-1 Sequence 1, Appli  
C 40 33 8.6 3282 1 US-08-276-852-154 Sequence 154, App  
41 33 8.6 3282 1 US-08-276-852-159 Sequence 169, App  
C 42 33 8.6 3282 1 US-08-899-575-154 Sequence 154, App  
43 33 8.6 3282 1 US-08-899-575-159 Sequence 169, App  
C 44 33 8.6 3282 1 US-08-899-575-154 Sequence 154, App  
45 33 8.6 3282 1 US-08-899-575-169 Sequence 169, App

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22133-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 INMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-F1s  
US-08-232-463-14  
Query Match 14.1%; Score 54.2; DB 1; Length 7218;

```

Best Local Similarity 2.9%; Pred.No. 6e-05;
Matches 11; Conservative 216; Mismatches 146; Indels 0; Gaps 0;

QY 10 GAGCAGGACGGGGCTGTGTGCCAGCTGCAGATCTGTGGCTGCTGGCAGATGCTGG 69
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1421 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 70 GGAGTACTCTGTGTGTGGGAGGAGAGCACTCTGCCACTCTACCCGTGAAGGCCCT 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1361 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 130 GCAGCGCAAGTTACAGAGAGGTGTGAGAATGAAGAGCGCTGGAAGGGCCACAGCCAT 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1301 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 190 GTTGTGTGTGAAGTGTGAGCAAGTGGCCCTGTGGAGTGAGGAAGGGCCGAGAACCT 249
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1241 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 250 CAGAGATGGGACAGATACATCTCTGAGCAGGAGGACCAAGTGTGAGTGCAGATCTG 309
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1181 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 310 TGCCTGGCCATGGCGACCGCGGAGTACTGTGTGTGCGGGCAGGAGGACCTC 369
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1121 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1062
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 370 AGCCAGCTCACCAT 384
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1061 AAGCTCCCTCGACCT 1047
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 2
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rih
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

```

APPLICANT: Buchbinder, Jenny  
TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS



FILE REFERENCE: PA-0041 US  
CURRENT APPLICATION NUMBER: US/09/976,594  
CURRENT FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: 60/240,409  
PRIOR FILING DATE: 2000-10-12  
NUMBER OF SEQ ID NOS: 1143  
SOFTWARE: PERL Program  
SEQ ID NO 941  
LENGTH: 7055  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. 6673549 346716.17  
US-09-976-594-941

Query Match 10.5%; Score 40.6; DB 4; Length 7055;  
Best Local Similarity 49.8%; Pred. No. 0.18;  
Matches 103; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 76 CTCCTGTGTGTGGGAGAGAGACCTCTGCCACTCTCACCCTGAAGCCCTGCCAGC 135  
Db 1767 CTTCTACCTGTGGACGAGGAGAGCAACTTCCCCACGCCACGAGCCAGACCTGTGTGC 1826

QY 136 CAAGTTCACAGAGGCTCTGAGGATGAGAGGCGTGAAGGGCCACACGCCATGTTGTG 195  
Db 1827 CAAGTTCAACAGCAACATGAGGCAATAGTACTTCTGGGACCCCGTCTATGAGCC 1886

QY 196 GTGTGAAGTGAAGGTGGCCCTCTGTGAGTGGAGGAAGGGCCCGAGAACCTCAGAGA 255  
Db 1887 AGCTTTTCATCTCCAGCACTTCGCGAGGAAGGTGAATATACAGATCAAGACTTCCGGGA 1946

QY 256 TGGGACAGATACATCTCTGAGGCAGGA 282  
Db 1947 GAAGAACTGAGTACATCGCGGCAGGA 1973

RESULT 5  
US-08-676-967-5  
Sequence 5, Application US/08676967  
Patent No. 5747317  
GENERAL INFORMATION:  
APPLICANT: COLLINS, KATHLEEN  
TITLE OF INVENTION: Human Telomerase  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Science & Technology Law Group  
STREET: 268 Bush Street, Suite 3200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/676,967  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UCB96-055  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415)343-4341  
TELEFAX: (415)343-4342  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double

TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-676-967-5

Query Match 10.1%; Score 39; DB 1; Length 2277;  
Best Local Similarity 45.0%; Pred. No. 0.37;  
Matches 147; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 14 CAGGACGGGGCTGTGTGGAGCTGCAGATCTGTGGCCCTGTGTGGCAGATGCTGGGGAG 73  
Db 406 CAGTTTCGGCGCGTCTGGAGGTGAACATCCCCCAAGCCGACGCAAGATGCGCGGC 465

QY 74 TACTCTGTGTGTGGGAGAGAGGACCTCTGCCACTCTCACCCTGAAGCCCTGCCA 133  
Db 466 TTCGGCTTGTGTGCTGAGTTCAGGAACCTGTGGAGGCCGCAAGCCCTGAAGGGCATGAAC 525

QY 134 GCCAAGTTTCACAGAGGCTCTGAGGAATGAAGAGGCGCTGGAAGGGGCCACAGCCATGTTG 193  
Db 526 ATGAGGAGATCAAGGGCCGACCCCTGGCCGTGGACTGGCCGTGCGCCAGGACAAGTAC 585

QY 194 TGTGTGAAGTGAAGGTGGCCCTCTGTGAGTGGAGGAAGGGCCCGAGAACCTCAGA 253  
Db 586 AAGGACACCCAGAGGCTGAGCGCCATCGCGAGGAGAAAGAGCCACGAGCAAGCACCAG 645

QY 254 GATGGGACAGATACATCTGAGGCGAGGAGGAGCCAGGTGTGAGCTGACATCTGTGGC 313  
Db 646 GAGAGCTGAAGAAGAGGGCCGCGAGGAGGAGGACATGGAGGAGGAGAGACGACGAC 705

QY 314 CTGGCCATGGCGGACGCCCGGGGAGTAC 340  
Db 706 GACGACGAGCAGCAGCAGCAGGAGGAG 732

RESULT 6  
US-08-676-974-5  
Sequence 5, Application US/08676974  
Patent No. 5770422  
GENERAL INFORMATION:  
APPLICANT: COLLINS, KATHLEEN  
TITLE OF INVENTION: Human Telomerase  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Science & Technology Law Group  
STREET: 268 Bush Street, Suite 3200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/676,974  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UCB96-055  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415)343-4341  
TELEFAX: (415)343-4342  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-676-974-5

Query Match 10.1%; Score 39; DB 1; Length 2277;  
Best Local Similarity 45.0%; Pred. No. 0.37; Indels 0; Gaps 0;  
Matches 147; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 14 CAGGACGGGGCTGTGTGCGAGCTCAGATCTGTGGCTGTGGCAGATGTGGGGAG 73  
DB 406 CAGTTGGGCGCGTGTGGAGTGAACATCCCCCGCAAGCCCGACGCAAGATGCGCGC 465

QY 74 TACTCTGTGTGTGGGAGGAGAGACCTCTGCCACTCTCACCCTGAAGGCCCTGCCA 133  
DB 406 CAGTTGGGCGCGTGTGGAGTGAACATCCCCCGCAAGCCCGACGCAAGATGCGCGC 465

QY 74 TACTCTGTGTGTGGGAGGAGAGACCTCTGCCACTCTCACCCTGAAGGCCCTGCCA 133  
DB 466 TTCCGCTTGTGTGCTCAAGAACCTCTGGAGCCCGCAAGCCCTGAAGGGCATGAAC 525

QY 134 GCCAAGTTTCAGAGGGTCTGAGGAATGAAGAGCCCTGTGGAAGGGCCACAGCCATGTTG 193  
DB 526 ATGAAGGAGATCAAGGCGCCGACCGTGTGGCTGTGGCCCAAGGACAAATAC 585

QY 194 TGGTGTGAACCTGAGCAAGTGGCCCTGTGGAGTGGAGGAGGGCCCGAGAACCTCAGA 253  
DB 586 AAGGACACCCAGAGCGTGAAGCCATCGGCGAGGAGAGAGCCACGAGAGCAAGCACAG 645

QY 254 GATGGGACAGATACATCTTGAAGAGAGCCCTGTGGAAGGGCCACAGCCATGTTG 193  
DB 646 GAGAGCTGTGAAGAGAGGCGCCGAGGAGGAGACATGAGGAGGAGGAGACGAGC 705

QY 314 CTGGCCATGGCGGACGCCGGGAGTAC 340  
DB 706 GACGACGACGACGACGAGGAGGAGC 732

## RESULT 7

US-09-098-487-5  
; Sequence 5, Application US/09098487  
; Patent No. 5917025  
; GENERAL INFORMATION:  
; APPLICANT: COLLINS, Kathleen  
; TITLE OF INVENTION: Human Telomerase  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Science & Technology Law Group  
; STREET: 268 Bush Street, Suite 3200  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/098,487  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman Ph.D., Richard A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: UCB96-055  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415)343-4341  
; TELEFAX: (415)343-4342  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2277 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; US-09-098-487-5

Query Match 10.1%; Score 39; DB 2; Length 2277;  
Best Local Similarity 45.0%; Pred. No. 0.37; Indels 0; Gaps 0;  
Matches 147; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 14 CAGGACGGGGCTGTGTGCGAGCTCAGATCTGTGGCTGTGGCAGATGTGGGGAG 73  
DB 406 CAGTTGGGCGCGTGTGGAGTGAACATCCCCCGCAAGCCCGACGCAAGATGCGCGC 465

QY 74 TACTCTGTGTGTGGGAGGAGAGACCTCTGCCACTCTCACCCTGAAGGCCCTGCCA 133  
DB 466 TTCCGCTTGTGTGCTCAAGAACCTCTGGAGCCCGCAAGCCCTGAAGGGCATGAAC 525

QY 134 GCCAAGTTTCAGAGGGTCTGAGGAATGAAGAGCCCTGTGGAAGGGCCACAGCCATGTTG 193  
DB 526 ATGAAGGAGATCAAGGCGCCGACCGTGTGGCTGTGGCCCAAGGACAAATAC 585

QY 194 TGGTGTGAACCTGAGCAAGTGGCCCTGTGGAGTGGAGGAGGGCCCGAGAACCTCAGA 253  
DB 586 AAGGACACCCAGAGCGTGAAGCCATCGGCGAGGAGAGAGCCACGAGAGCAAGCACAG 645

QY 254 GATGGGACAGATACATCTTGAAGAGAGCCCTGTGGAAGGGCCCGAGAACCTCAGA 253  
DB 646 GAGAGCTGTGAAGAGAGGCGCCGAGGAGGAGACATGAGGAGGAGGAGACGAGC 705

QY 314 CTGGCCATGGCGGACGCCGGGAGTAC 340  
DB 706 GACGACGACGACGACGAGGAGGAGC 732

## RESULT 8

US-09-620-312D-709/c  
; Sequence 709, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunqing  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: NO. 6569662e1 Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: PL-FL\_genes Version 1.0  
; SEQ ID NO 709  
; LENGTH: 707  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (339)...(698)  
; US-09-620-312D-709

Query Match 9.8%; Score 37.6; DB 4; Length 707;  
Best Local Similarity 44.8%; Pred. No. 0.66;  
Matches 145; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 11 AGGCAGGAGCGGGCTGTGTGGAGCTGCAGATCTGTGGCTGTGGCAGATGTGGG 70



TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11461 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-669-161A-29

Query Match 9.7%; Score 37.2; DB 3; Length 11461;  
Best Local Similarity 51.2%; Pred. No. 1.5;  
Matches 87; Conservative 0; Mismatches 83; Indels 0; Gaps 0;  
QY 186 CCATGTTGTGTGTAAGTCAAGTGGCCCTGTGGAGTGGAGGAGGGCCCGAGA 245  
DB 2772 CCAGCCGAGCAGTTCAAGTGAAGTGAACAGCAACACCTGAAGAGGGGAACGAC 2831  
QY 246 ACCTCAGAGATGGGACAGATACATCTCTGAGGAGGAGGGACCAAGTGTGAGCTGCAGA 305  
DB 2832 AACTCAACGTGAGATCTCTGAGTGTCTCAGGAGGGAGAGGATCAGGAGCATCTGCAG 2891  
QY 306 TCTGTGCTGCTGCGCATGCGGACCGCGGGAGTACTTGTGTGTGCGGG 355  
DB 2892 TCAGGTGAGGAGGAGGAGGAGCAACCTCTGCTGTGTGTGCTGTGTG 2941

RESULT 12  
US-09-687-050-5  
Sequence 5, Application US/09687050  
Patent No. 6632637

GENERAL INFORMATION:  
APPLICANT: McGrew, Jeffrey T.  
TITLE OF INVENTION: VECTORS AND METHODS FOR RECOMBINANT PROTEIN EXPRESSION  
FILE REFERENCE: 2902-A  
CURRENT APPLICATION NUMBER: US/09/687,050  
PRIOR FILING DATE: 2000-10-12  
PRIOR APPLICATION NUMBER: 60/159,177  
PRIOR FILING DATE: 1999-10-13  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 5

LENGTH: 2478  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(2478)  
NAME/KEY: mat\_peptide  
LOCATION: (76)...()  
NAME/KEY: sig\_peptide  
LOCATION: (1)...(75)  
US-09-687-050-5

Query Match 9.6%; Score 37; DB 4; Length 2478;  
Best Local Similarity 49.2%; Pred. No. 1.2;  
Matches 97; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
QY 132 CAGCAAGTTCACAGAGGGTCTGAGGAATGAGAGGCGGTGGAGGGCCAGAGCCATGT 191  
DB 1016 CAGCATGTCGCCAGTGGAGATCAGCAAGACATCTCTGGCCAGAGATCAGCGTGG 1075  
QY 192 TGTGTTGTAAGTCAAGTGGCCCTGTGGAGTGGAGGAGGGCCCGAGAACTCA 251  
DB 1076 TCGATGTGTGGAGTTGTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1135  
QY 252 GAGATGGGACAGATACATCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 311  
DB 1136 AAGAAAAGGAGGCTTCTGTGCATCGCCCTGAGAGCAGAGGGATGACTTCCAGAGGAA 1195  
QY 312 GCCTGGCCATGGCGGAC 328

DB 1196 GGGAGGGCATTGTGGCC 1212

RESULT 13  
US-09-921-667-15  
Sequence 15, Application US/09921667  
Patent No. 6652854

GENERAL INFORMATION:  
APPLICANT: Mohler, Kendall M.  
APPLICANT: Barone, Dauphine S.  
APPLICANT: Peschon, Jacques J.  
APPLICANT: Kennedy, Mary K.  
APPLICANT: Pluennske, John D.  
TITLE OF INVENTION: METHODS FOR TREATING AUTOIMMUNE AND CHRONIC INFLAMMATORY CONDITIONS  
FILE REFERENCE: 2959-A  
CURRENT APPLICATION NUMBER: US/09/921,667  
CURRENT FILING DATE: 2001-08-03  
PRIOR APPLICATION NUMBER: 60/224,079  
PRIOR FILING DATE: 2000-08-08  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 15

LENGTH: 2478  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(2475)  
OTHER INFORMATION:  
NAME/KEY: mat\_peptide  
LOCATION: (76)...()  
OTHER INFORMATION:  
US-09-921-667-15

Query Match 9.6%; Score 37; DB 4; Length 2478;  
Best Local Similarity 49.2%; Pred. No. 1.2;  
Matches 97; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
QY 132 CAGCAAGTTCACAGAGGGTCTGAGGAATGAGAGGCGGTGGAGGGCCAGAGCCATGT 191  
DB 1016 CAGCATGTCGCCAGTGGAGATCAGCAAGACATCTCTGGCCAGAGATCAGCGTGG 1075  
QY 192 TGTGTTGTAAGTCAAGTGGCCCTGTGGAGTGGAGGAGGGCCCGAGAACTCA 251  
DB 1076 TCGATGTGTGGAGTTGTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1135  
QY 252 GAGATGGGACAGATACATCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 311  
DB 1136 AAGAAAAGGAGGCTTCTGTGCATCGCCCTGAGAGCAGAGGGATGACTTCCAGAGGAA 1195  
QY 312 GCCTGGCCATGGCGGAC 328  
DB 1196 GGGAGGGCATTGTGGCC 1212

RESULT 14  
US-09-016-434-1319  
Sequence 1319, Application US/09016434  
Patent No. 6500938

GENERAL INFORMATION:  
APPLICANT: Janice Au-Young  
APPLICANT: Jeffrey J. Seilhamer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
NUMBER OF SEQUENCES: 1490  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,434
FILING DATE: HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0002 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1319:
SEQUENCE CHARACTERISTICS:
LENGTH: 3597 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: g33833
US-09-016-434-1319

Query Match          9.6%; Score 37; DB 4; Length 3597;
Best Local Similarity 49.2%; Pred. No. 1.3; Indels 0; Gaps 0;
Matches 97; Conservative 0; Mismatches 100;

QY 132 CAGCCAAAGTTCACAGAGGGTCTGAGGAATGAAGAGGCCGCTGGAAGGGCCGACAGCCATGT 191
Db 1191 CAGCATGGTCCAGTGGAGATCAGCAAGACAGTCTCTGGCCAGAGAGCATCAGCGTGG 1250

QY 192 TGTGGTGAACACTGAGCAAGGTGGCCCTCTGGAGTGGAGGAGGGCCCGAGAACCTCA 251
Db 1251 TCGCATGTGTGGAGTTGTTTGGAGCCCGCTGGAGTGTGAGGAGGAGGAGGTAGAGG 1310

QY 252 GAGATGGGGACAGATACATCTCTGAGGAGGAGGAGGAGGAGGAGTGTGAGTGTGAGTGTG 311
Db 1311 AAGAAAAGGAGCTTCTGTGCATCGCTGAGAGCAGCAGGAGTACTTCCAGAGGGAA 1370

QY 312 GCCTGGCCATGGCGAC 328
Db 1371 GGGAGGGCATTGTGGCC 1387

Search completed: September 19, 2004, 22:12:36
Job time : 35.5752 secs
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,434
FILING DATE: HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0002 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1319:
SEQUENCE CHARACTERISTICS:
LENGTH: 3597 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: g33833
US-09-016-434-1319

Query Match          9.6%; Score 37; DB 4; Length 3597;
Best Local Similarity 49.2%; Pred. No. 1.3; Indels 0; Gaps 0;
Matches 97; Conservative 0; Mismatches 100;

QY 132 CAGCCAAAGTTCACAGAGGGTCTGAGGAATGAAGAGGCCGCTGGAAGGGCCGACAGCCATGT 191
Db 1191 CAGCATGGTCCAGTGGAGATCAGCAAGACAGTCTCTGGCCAGAGAGCATCAGCGTGG 1250

QY 192 TGTGGTGAACACTGAGCAAGGTGGCCCTCTGGAGTGGAGGAGGGCCCGAGAACCTCA 251
Db 1251 TCGCATGTGTGGAGTTGTTTGGAGCCCGCTGGAGTGTGAGGAGGAGGAGGTAGAGG 1310

QY 252 GAGATGGGGACAGATACATCTCTGAGGAGGAGGAGGAGGAGGAGTGTGAGTGTGAGTGTG 311
Db 1311 AAGAAAAGGAGCTTCTGTGCATCGCTGAGAGCAGCAGGAGTACTTCCAGAGGGAA 1370

QY 312 GCCTGGCCATGGCGAC 328
Db 1371 GGGAGGGCATTGTGGCC 1387

RESULT 15
US-09-023-655-1314
Sequence 1314, Application US/09023655
Patent No. 6607879
GENERAL INFORMATION:
APPLICANT: Cocks, Benjamin G.
APPLICANT: Susan G. Stuart
APPLICANT: Jeffrey J. Seilhamer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSES: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 12:53:28 ; Search time 158.742 Seconds  
(without alignments)  
10303.209 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_10286\_10670

Perfect score: 385  
Sequence: 1 acatagctgagcaggagc.....cctcagccacgctcaccatc 385

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 337363 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq 29Jan04.\*  
1: geneseqn1980s.\*  
2: geneseqn1990s.\*  
3: geneseqn2000s.\*  
4: geneseqn2001as.\*  
5: geneseqn2001bs.\*  
6: geneseqn2002s.\*  
7: geneseqn2003as.\*  
8: geneseqn2003bs.\*  
9: geneseqn2003cs.\*  
10: geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	385	100.0	7564	7 ABZ24581	Abz24581 Human cel
2	385	100.0	14061	6 ABV99363	Abv99363 Human NOV
3	385	100.0	14109	6 ABV99362	Abv99362 Human NOV
4	385	100.0	24120	7 ABX11642	Abx11642 Human ser
5	253.6	65.9	352	5 AAF64470	Aaf64470 Novel hum
6	202	52.5	898	6 ABL58139	AbL58139 Human pho
7	195.4	50.8	642	6 ABO61180	AbG61180 Obscurin
8	195.4	50.8	707	6 ABO61180	AbG61180 Obscurin
9	190.6	49.5	2155	6 ABX99965	Abx99965 DNA encod
10	71.6	18.6	1005	6 ABN21414	Abn21414 Human ORF
11	47.6	12.4	2737	4 AAI59251	Aai59251 Human pol
12	47.6	12.4	3999	4 AAI61037	Aai61037 Human pol
13	47.6	12.4	5382	9 ADI14722	Adi14722 Human src
14	46.8	12.2	81940	4 AAS05390	Aas05390 Human tit
15	46.8	12.2	81940	6 ABK64829	Abk64829 Human ben
16	46.8	12.2	93801	8 ABX13540	Abx13540 Human RGS
17	46.2	12.0	60	6 AEN40768	Abn40768 Human spl
18	43.4	11.3	1845	5 AAF24162	Aaf24162 Human sec
19	42.8	11.1	2254	3 AAC93403	Aac93403 Human sec
20	42.8	11.1	2254	5 AAF24183	Aaf24183 Human sec
21	41.8	10.9	716	6 ABK35726	Abk35726 CDNA sequ
22	41.8	10.9	1948	2 ABV74346	Abv74346 Human IL-
23	41.8	10.9	1949	2 ABV74347	Abv74347 Human IL-

24	41.8	10.9	2170	6 ABQ54970	Abq54970 Human ova
25	40.6	10.5	3761	4 ABL05157	AbL05157 Drosophila
26	40.6	10.5	6069	9 ADD18703	Add18703 Human dis
27	40.6	10.5	7066	10 ADE77197	Ade77197 Human cdn
28	39.4	10.2	1359	2 AAZ17254	Aaz17254 Human gen
29	39	10.1	2277	2 AAV13836	Aav13836 Homo sapi
30	39	10.1	2277	2 AAV05372	Aav05372 Human tel
31	38.6	10.0	5452	9 ADC86736	Adc86736 Human gpc
32	38.2	9.9	1251	9 ADC30509	Adc30509 Human nov
33	38.2	9.9	2488	4 AAH18453	Aah18453 Human cdn
34	38.2	9.9	67251	9 ADC26995	Adc26995 Sorangium
35	38	9.9	8047	4 ABA07219	AbA07219 Human pan
36	38	9.9	8047	4 AAK89856	Aak89856 Human dig
37	37.8	9.8	17127	7 ADA98892	Ada98892 Human sec
38	37.8	9.8	17127	7 ADA44493	Ada44493 Human sec
39	37.8	9.8	17127	9 ADC20927	Adc20927 Human sec
40	37.6	9.8	672	6 ABQ98892	AbQ98892 Human ORF
41	37.6	9.8	707	4 AAI58818	Aai58818 Human pol
42	37.6	9.8	707	8 ADB48799	AdB48799 Novel hum
43	37.6	9.8	1068	7 ADA70882	Ada70882 Rice gene
44	37.6	9.8	2121	4 AAK94576	Aak94576 Human ful
45	37.6	9.8	2437	6 AEN85621	Abn85621 Human cdn

#### ALIGNMENTS

RESULT 1  
ABZ24581  
ID ABZ24581 standard; cdna; 7564 BP.

XX AC ABZ24581;

DT 31-MAR-2003 (first entry)

XX DE Human cell adhesion and extracellular matrix protein 4 cDNA.

XX KW Cell adhesion and extracellular matrix protein 4; CADECM-4; human;  
XX KW anti-HIV; virucide; anti-allergic; anti-inflammatory; antianemic;  
XX KW antiparkinsonian; nootropic; anticonvulsant; antinfertility;  
XX KW antitartaroclastic; antiaesthetic; immunosuppressive; antithyroid;  
XX KW cytostatic; hepatotropic; dermatological; antidiabetic; nephrotropic;  
XX KW antigout; thyromimetic; neuroprotective; osteopathic; antiarthritic;  
XX KW antiparasitic; antihelminthic; antiprosoritic; uropathic; ophthalmological;  
XX KW antirheumatic; haemostatic; antibacterial; protozoacide; fungicide;  
XX KW gynaecological; titin; gene therapy; gene; ss.

OS Homo sapiens.

XX FH Key Location/Qualifiers  
XX CDS 241..7227  
XX /\*tag= a  
XX /product= "Human CADECM-4"

XX DN WC200288322-A2.

XX PD 07-NOV-2002.

XX PF 01-MAY-2002; 2002WO-US013874.

XX PR 02-MAY-2001; 2001US-02889290P.

XX PR 21-MAY-2001; 2001US-0292468P.

XX PR 15-JUN-2001; 2001US-0299616P.

XX PR 28-JUN-2001; 2001US-0301672P.

XX PR 04-JAN-2002; 2002US-0345008P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Yue H, Lee EA, Duggan BM, Thangavelu K, Honchell CD, Ding Li, Tran UK;  
XX PI Hillman JL, Baughn MR, Kallick DA, Lee S, Warren BA, Xu Y, AR;  
XX PI Lal PG, Thornton M, Hafalia AUA, Yao MG, Nguyen DB, Gandhi AR;  
XX PI Khan FA, Walhia NK, Griffin JA, Chinn AM, Elliott VS, Ramkumar J;  
XX PI Arvizu CS, Forsythe LJ;





PR	12-SEP-2001;	2001US-0318770P.	
PR	27-SEP-2001;	2001US-0325430P.	
PR	27-SEP-2001;	2001US-0325681P.	
PR	18-OCT-2001;	2001US-0330380P.	
PR	31-OCT-2001;	2001US-0335301P.	
PR	14-NOV-2001;	2001US-0332172P.	
PR	14-NOV-2001;	2001US-0332271P.	
PR	14-NOV-2001;	2001US-0332272P.	
PR	14-NOV-2001;	2001US-0333184P.	
PR	14-NOV-2001;	2001US-0333272P.	
PR	21-NOV-2001;	2001US-0332094P.	
PR	03-DEC-2001;	2001US-0337426P.	
PR	03-DEC-2001;	2001US-0338092P.	
PR	04-DEC-2001;	2001US-0337185P.	
PR	03-JAN-2002;	2002US-0345705P.	
PR	08-MAR-2002;	2002US-00093463.	
XX			
XX		(CURA-) CURAGEN CORP.	
XX			
PI	Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;		
PI	Bolog FL, Li L, Zernhesen BD, Ichernev VT, Gangelli EA, Vernet CAM;		
PI	Pena CE, Burgess CE, Liu X, Spytek KA, Gorman L, Spaderna SK;		
PI	Voss EZ, Malyankar UM, Anderson DW, Patturajan M, Miller CE;		
PI	Tauber RJ, Padigaru M, Shenoy SG, Kekuda R, Gusev VY, Pochart PF;		
PI	Zhong M;		
XX			
XX	WPI: 2002-732824/79.		
DR	P-PSDB; ABP70085.		
XX			
PT	New NOVX polypeptides and polynucleotides, useful for preventing,		
PT	diagnosing or treating NOVX-associated disorders e.g. diabetes, cancer,		
PT	Alzheimer's disease, dyslipidemias, obesity, immune or hematopoietic		
PT	disorders, and asthma.		
XX			
XX	Claim 16; Page 138-142; 619pp; English.		
XX			
CC	The present invention relates to new isolated proteins (NOVX) and their		
CC	coding sequences (ABV99327-ABV99595 and ABP70049-ABP70149), where X is		
CC	any number from 1 to 48. The NOVX proteins and coding sequences are		
CC	useful in the manufacture of a medicament for treating a syndrome		
CC	associated with a human disease, preferably a NOVX-associated disorder.		
CC	The NOVX coding sequences and proteins are useful for treating, diabetes,		
CC	preventing or diagnosing diseases such as metabolic disorders, diabetes,		
CC	obesity, infectious disease, anorexia, cancer-associated cachexia,		
CC	cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's		
CC	disease, immune disorders, hematopoietic disorders, cardiovascular		
CC	disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic		
CC	disturbances associated with obesity, metabolic syndrome X or wasting		
CC	disorders associated with chronic diseases or various cancers. The NOVX		
CC	coding sequences and proteins may also be used as targets for the		
CC	identification of small molecules that modulate or inhibit e.g.		
CC	neurogenesis, cell differentiation, cell proliferation, haematopoiesis,		
CC	wound healing and angiogenesis. In gene therapy, in generation of		
CC	antibodies that bind immunospecifically to NOVX substances for use in		
CC	therapeutic or diagnostic methods		
XX			
XX	Sequence 14061 BP; 2854 A; 4029 C; 4839 G; 2339 T; 0 U; 0 Other;		
QQ			
QQ	Query Match 100.0%; Score 385; DB 6; Length 14061;		
QQ	Best Local Similarity 100.0%; Pred No. 4.2e-86;		
QQ	Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 ACATAGCCTGAGCAGGAGCGGGGCTGTGTGCGAGCTGCAGATCTGTGGCTGTGGC 60		
DB	2447 ACATAGCCTGAGCAGGAGCGGGGCTGTGTGCGAGCTGCAGATCTGTGGCTGTGGC 2506		
QY	61 AGATGCTGGGAGTACTCTGTGTGTGTGTGGGAGGAGGACCTCTGCCACTCTACCGT 120		
DB	2507 AGATGCTGGGAGTACTCTGTGTGTGTGTGGGAGGAGGAGGACCTCTGCCACTCTACCGT 2566		
QY	121 GAAGGCCCTCCAGCCAAAGTTTCACAGAGGGTCTGTGAGGAATGAAGAGCCCTGGAAGGGC 180		
DB	2567 GAAGGCCCTCCAGCCAAAGTTTCACAGAGGGTCTGTGAGGAATGAAGAGCCCTGGAAGGGC 2626		



XX	15-FEB-2002; 2002US-00077130.
PF	
XX	
PR	15-FEB-2001; 2001US-0269201F.
XX	(MILL-) MILLENNIUM PHARM INC.
XX	Kapeller-Libermann R, Acton SL;
PI	WPI; 2003-398729/29.
DR	P-PSDB; ABG76187.
XX	
XX	Novel isolated human protein kinase, designated 59079 or 12599 polypeptide, useful as diagnostic and therapeutic agents for preventing cardiovascular diseases, proliferative disorders, and protein kinase disorders.
PT	
PT	
PT	
XX	Claim 2; Page 58-84; 119pp; English.
PS	The invention relates to an isolated human serine/threonine or protein kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule comprising at least 85% identity to the nucleic acids appearing as AX111641 and AX111642 or their complement, a naturally occurring variant of the kinases or their fragments. Also included are a non-human host cell containing the nucleic acids, an antibody specific for the proteins, identifying a compound which binds to the kinase (by contacting the kinase or a cell expressing the kinase with a test compound and determining whether the kinase binds to the test compound) and modulating the activity of kinase using the identified compound. The kinases and their encoding nucleic acids are useful as diagnostic and therapeutic agents for preventing a disease or condition associated with an aberrant or unwanted 59079 or 12599 activity in a subject, including cardiovascular diseases such as heart failure, and myocardial infarction; disorders involving blood vessels such as atherosclerosis, and Kaposi's sarcoma; blood platelets disorder such as thrombocytopoenia, leukaemia, Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders such as cancer; and protein kinase disorders such as autoimmune disorders, diabetes mellitus, psoriasis, inflammatory bowel disease, rheumatoid arthritis, and multiple sclerosis (many examples of diseases and disorders are included in the specification). The kinases, their encoding nucleic acids and antibodies are useful in screening assays, detection assays (e.g. forensic biology), and predictive medicine (e.g. diagnostic assays, prognostic assays, and monitoring clinical trials and pharmacogenomics). The kinases and their encoding nucleic acids are useful as query sequences to perform a search against public databases to identify other family members or related sequences. The present sequence encodes the kinase 12599
XX	
SQ	Sequence 24120 BP; 4719 A; 7265 C; 8278 G; 3858 T; 0 U; 0 Other;
	Query Match 100.0%; Score 385; DB 7; Length 24120;
	Best Local Similarity 100.0%; Pred. No. 4.7e-86;
	Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 ACATAGCCTGAGCGAGGACGGGCGTGTGCCAGCTGCAGATCTGTGGCCCTGGCTGTGGC 60
Db	10286 ACATAGCCTGAGCGAGGACGGGCGTGTGCCAGCTGCAGATCTGTGGCCCTGGCTGTGGC 10345
QY	61 AGATGTGGGGAGTACTCTGTGTGTGTGGGGAGGAGGACCTCTGCCACTCTCACCGT 120
Db	10346 AGATGTGGGGAGTACTCTGTGTGTGTGGGGAGGAGGACCTCTGCCACTCTCACCGT 10405
QY	121 GAAGGCCCTGCCAGCCAACTTCACAGAGGCTCTGAGGAATGAGAGGCCCTGGAGGGGC 180
Db	10406 GAAGGCCCTGCCAGCCAACTTCACAGAGGCTCTGAGGAATGAGAGGCCCTGGAGGGGC 10465
QY	181 CACAGCCCATGTTGTGGTGTAACCTGAGCAAGTGGCCCCCTGTGGAGTGGAGAGGGGCC 240
Db	10466 CACAGCCATGTTGTGGTGTAACCTGAGCAAGTGGCCCCCTGTGGAGTGGAGAGGGGCC 10525
QY	241 CGAAGACCTCAGAGATGGGGAAGATACATCTCTGAGGAGGAGGGGAACAGGTGTAGCT 300
Db	10526 CGAAGACCTCAGAGATGGGGAAGATACATCTCTGAGGAGGAGGGGAACAGGTGTAGCT 10585

Qy	301	GCAGATCTGTGCGCTGGCCATCGCGGAGCGCGGGGAGTACTGTGTGTGTCGGGCAGGA	360
Db	10586	GCAGATCTGTGCGCTGGCCATCGCGGAGCGCGGGGAGTACTGTGTGTGTCGGGCAGGA	10645
Qy	361	GAGGACCTCAGGCACGCTCACCATC	385
Db	10646	GAGGACCTCAGGCACGCTCACCATC	10670

## RESULT 5

AA64470/C  
ID  
AAF64470 standard; cDNA; 352 BP.  
XX  
XX  
AC  
AC  
AAF64470;  
XX  
XX  
09-APR-2001 (first entry)  
DT  
DT  
XX  
XX  
DE Novel human polynucleotide, SEQ ID NO: 226.

Human; cytostatic; gene therapy; colon cancer; prostate cancer;  
 breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

XXXXXX

XX  
ED  
11-JAN-2001.

PF 30-JUN-2000;  
v.v.

PR 02-JUL-1999;

PK 02-JUL-1999;  
XX

PA (CHIR ) CHIR ) HVSE - ) HVSE - )

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

PI Kassam A, R

PI Dymnac R,  
PT Ieshkowitz

XX

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

PT Library of p  
PT mammalian ce

PT prostate, co  
vv

PS Claim 9; pag

XX The present

polynucleoti

CC detect colon

CC produce p  
CC of the polyn

CC the polynuc  
CC or antisense  
CC

their gene p

CC  
CC  
CC

preventive i  
antibodies a

CC treat the ca

3  
3  
4  
2  
3  
4  
2  
2  
1  
3  
00 XX  
00 XX

sequence 352

Query Match

Matches 280;

Ov 62 GAT

Db 342 GACCTGGGGAGTACCTGTCATGTGCGGGAAGAGAGGACCTCAGCCATGCTCACCGTC 283  
QY 122 AAGCCCTGCCAGCAAGTTTACAGAGGGTCTGAGGAATGAAGAGCCGTGGAAGGGCC 181  
Db 282 AGGCCCATGCTTCCAAAGTTTCATAGAGGTCTGAGGAATGAAGAGGCCACAGAAAGGGAC 223  
QY 182 ACAGCCATGTTGTGGTGTGAACCTGAGCAAGGTGCGCCCTGTGAGTGGAGGAAGGGGCC 241  
Db 222 ACGGCCACGCTGTGGTGTGAGCTGAGCAAGCGGCCACCGGTGAGTGGAGGAAGGGGCAT 163  
QY 242 GAGAACCTCAGAGATGGGGACAGATACATCTGAGGAGAGAGGGGACAGGTGTGAGCTG 301  
Db 162 GAGACCTCAGAGATGGGGACAGACACAGCTCAGGAGGATGGGTCCAGGTGTGAGCTG 103  
QY 302 CAGATCTGTGCGCTGGCCATGGCGGAGCGCGGGAGTACTTGTGTGTCGGGAGGAG 361  
Db 102 CAGATCGGTGCGCTGGCTGTGGTGGATGCCGGGAGTACTTGTGTGTCGGGAGGAG 43  
QY 362 AGGACCTCAGCCACGCTCACCATC 385  
Db 42 AGGACCTCAGCCACACTCACTGTC 19

RESULT 6  
ABL58139  
ID ABL58139 standard; cDNA; 898 BP.  
AC ABL58139;  
XX  
DT 30-JUL-2002 (first entry)  
XX  
DE Human phosphorylcholine/ethanolamine transferase 15.29 coding sequence.  
XX  
KW Human; phosphorylcholine/ethanolamine transferase 15.29; enzyme;  
KW phosphorylcholine; ethanolamine; transferase; tumour; inflammation;  
KW cytosolic; anti-inflammatory; gene; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
Key Location/Qualifiers  
FH 411..830  
FT CDS /\*tag= a  
FT /product= "Human phosphorylcholine/ethanolamine  
FT transferase 15.29"  
XX  
CN1331315-A.  
XX  
PD 16-JAN-2002.  
XX  
PF 30-JUN-2000; 2000CN-00116906.  
XX  
XX 30-JUN-2000; 2000CN-00116906.  
PR (BODE-) BODE GENE DEV CO LTD SHANGHAI.  
XX  
FA Mao Y, Xie Y;  
XX  
PI WPI; 2002-292894/34.  
XX  
DR P-PSDB; ABB83109.  
XX  
PT Human phosphorylcholine/ethanolamine transferase 15.29 and encoding  
PT polynucleotide useful for treating tumors and inflammation.  
XX  
PS Claim 6; Page 27 (Disclosure); 34pp; Chinese.  
XX  
XX The present sequence is the coding sequence for human phosphorylcholine/  
CC ethanolamine transferase 15.29. The enzyme and its coding sequence are  
CC useful for treating diseases such as tumours and inflammation  
XX  
SQ Sequence 898 BP; 169 A; 245 C; 239 G; 245 T; 0 U; 0 Other;  
Query Match 52.5%; Score 202; DB 6; Length 898;

Best Local Similarity 83.6%; Pred. No. 9.3e-41;  
Matches 229; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
QY 112 TCTACCTGTAAGGCCCTGCCAGCAAGTTTACAGAGGGTCTGAGGAATGAAGAGCGGT 171  
Db 335 TCTGACCTCCCGAGCCCTGCCGCAAGTTTACAAAGGGTCTGAGGAATGAAGAGGCCAC 394  
QY 172 GGAAGGGCCACAGCCATGTTGTGGTGTGAACCTGAGCAAGTGGCCCTCTTGGAGTGGAG 231  
Db 395 AGAAGGGCCACGACTAATGTTGAGTGTGAGTGTGAGCAAGTGGCCCTCTTGGAGTGGAG 454  
QY 232 GAAGGGCCCGAGAACCTCAGAGATGGGGACAGATACATCTGAGGAGGAGGGGACAG 291  
Db 455 GAAGGGACCTGAAACCTCAGAGATGGGGACAGATACAACTGAGGAGGATGGGACAG 514  
QY 292 GTGTGAGTGTGAGATCTGTGCGCTGGCCATGGCGGAGCGCGGGAGTACTTGTGTGTG 351  
Db 515 ATGTGAGTGTGAGATCTGTGCGCTGTGCGTGGCAGACACTGGGGAGTACTCAATGTATG 574  
QY 352 CGGGCAGAGAGGAGCCTCAGCAAGTGGCCATGGCGGAGCGCGGGAGTACTTGTGTGTG 385  
Db 575 TGGTCAGGAGAAGACGTCGGCCACTTCACTGTC 608  
RESULT 7  
ABO61169  
ID ABO61169 standard; cDNA; 642 BP.  
XX ABO61169;  
XX  
DT 26-FEB-2003 (first entry)  
XX  
DE Obscure OBSCN gene encoding sequence.  
XX  
KW Neuroprotective; immunomodulator; cancer; cytostatic; anti-inflammatory;  
KW gene therapy; nutritional supplement; wound; burn; ulcer;  
KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
KW autoimmune disorder; inflammation; vulnery; gene; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200231111-A2.  
XX  
PD 18-APR-2002.  
XX  
PF 11-OCT-2001; 2001WO-US027760.  
XX  
PR 12-OCT-2000; 2000US-00687527.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX  
DR WPI; 2002-426278/45.  
DR N-PSDB; ABP43925.  
XX  
PT New polypeptides and their encoded proteins, useful as nutritional  
PT sources or supplements, or in gene therapy, particularly for treating  
PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or  
PT inflammation.  
XX  
PS Claim 1; SEQ ID # 382; 357pp + Sequence Listing; English.  
XX  
CC The invention relates to 446 newly isolated polynucleotide sequences. The  
CC activity of polynucleotides of the invention may be described as,  
CC vulnery, neuroprotective, immunomodulator, cytostatic and anti-  
CC inflammatory. Compositions comprising nucleic acids of the invention are  
CC useful for treating a mammalian subject, or as nutritional sources or  
CC supplements. These are useful in gene therapy, particularly for treating  
CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,  
CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or  
CC inflammation. The nucleic acids and polypeptides are also useful in

CC diagnostic and research methods. The sequences given in records ABQ60788-  
CC ABQ61233 represent polynucleotides of the invention. NOTE: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 642 BP; 147 A; 169 C; 205 G; 121 T; 0 U; 0 Other;

Query Match 50.8%; Score 195.4; DB 6; Length 642;  
Best Local Similarity 84.3%; Pred. No. 3.8e-39;  
Matches 220; Conservative 0; Mismatches 41; Indels 0; Gaps 0;  
QY 125 GCCTGCCAGCAAGTTCACAGAGGGTCTGAGGAATCAAGAGGCCGTGGAGGGGCCACA 184  
DB 227 GCTTACCATCAATTCACAGAGGGTCTGAGGAACGAGAGGCCACAGAGGGGCAACA 286  
QY 185 GCCATGTTGTGTGAACCTGAGCAAGTGGCCCTCTGTGAGTGAGGAAGGCCCGCAG 244  
DB 287 GCCGTGTGCGTGTGAGCTGAGCAAGATGCCCTGAGTGTGAGTGTGAGAGGGCGATGAG 346  
QY 245 AACCTCAGAGATGGGACAGATACATCTGAGGCAGGAGGGACCCAGTGTGAGCTGCAG 304  
DB 347 ACCCTCAGAGATGGAGACAGACAGAGCTGAGGCAGGACGGGGCCAGTGTGAGCTGCAG 406  
QY 305 ATCTGTGCGCTGGCCATGGCGAGCGCCGGGAGTACTTGTGTGTGTCGGCGCAGGAGG 364  
DB 407 ATCCGCGCTCTGTCGAGAGGACGCTGGGAGTACCTGTGATGTGCGGGAGGAGG 466  
QY 365 ACCTCAGCCAGCTCACCATC 385  
DB 467 ACCTCAGCCATGCTCACCGTC 487

RESULT 8  
ABQ61180  
ID ABQ61180 standard; cDNA; 707 BP.  
XX  
AC ABQ61180;  
XX  
DT 26-FEB-2003 (first entry)  
XX  
DE Obscurin OBSCN gene #2 encoding sequence.

XX Neuroprotective; immunomodulator; cancer; cytostatic; anti-inflammatory;  
XX gene therapy; nutritional supplement; wound; burn; ulcer;  
KW Alzheimer's disease; Huntington's disease; amyotrophic lateral sclerosis;  
KW autoimmune disorder; inflammation; vulnery; gene; ss.  
XX

OS Homo sapiens.  
XX  
PN WO200231111-A2.  
XX  
PD 18-APR-2002.  
XX  
PF 11-OCT-2001; 2001WO-US027760.  
XX  
PR 12-OCT-2000; 2000US-00687527.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX  
DR WPI; 2002-426278/45.  
DR N-PSDB; ABP43936.  
XX

XX New polypeptides and their encoded proteins, useful as nutritional  
PT sources or supplements, or in gene therapy, particularly for treating  
PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or  
PT inflammation.

PS Claim 1; SEQ ID # 393; 357bp + Sequence Listing; English.  
XX

CC The invention relates to 446 newly isolated polynucleotide sequences. The  
CC activity of polynucleotides of the invention may be described as,  
CC vulnery, neuroprotective, immunomodulator, cytostatic and anti-  
CC inflammatory. Compositions comprising nucleic acids of the invention are  
CC useful for treating a mammalian subject, or as nutritional sources or  
CC supplements. These are useful in gene therapy, particularly for treating  
CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,  
CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or  
CC inflammation. The nucleic acids and polypeptides are also useful in  
CC diagnostic and research methods. The sequences given in records ABQ60788-  
CC ABQ61233 represent polynucleotides of the invention. NOTE: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 707 BP; 154 A; 198 C; 213 G; 142 T; 0 U; 0 Other;

Query Match 50.8%; Score 195.4; DB 6; Length 707;  
Best Local Similarity 84.3%; Pred. No. 3.9e-39;  
Matches 220; Conservative 0; Mismatches 41; Indels 0; Gaps 0;  
QY 125 GCCTGCCAGCAAGTTCACAGAGGGTCTGAGGAATCAAGAGGCCGTGGAGGGGCCACA 184  
DB 227 GCTTACCATCAATTCACAGAGGGTCTGAGGAACGAGAGGCCACAGAGGGGCAACA 286  
QY 185 GCCATGTTGTGTGAACCTGAGCAAGTGGCCCTCTGTGAGTGAGGAAGGCCCGCAG 244  
DB 287 GCCGTGTGCGTGTGAGCTGAGCAAGATGCCCTGAGTGTGAGTGTGAGAGGGCGATGAG 346  
QY 245 AACCTCAGAGATGGGACAGATACATCTGAGGCAGGAGGGACCCAGTGTGAGCTGCAG 304  
DB 347 ACCCTCAGAGATGGAGACAGACAGAGCTGAGGCAGGACGGGGCCAGTGTGAGCTGCAG 406  
QY 305 ATCTGTGCGCTGGCCATGGCGAGCGCCGGGAGTACTTGTGTGTGTCGGCGCAGGAGG 364  
DB 407 ATCCGCGCTCTGTCGAGAGGACGCTGGGAGTACCTGTGATGTGCGGGAGGAGG 466  
QY 365 ACCTCAGCCAGCTCACCATC 385  
DB 467 ACCTCAGCCATGCTCACCGTC 487

RESULT 9  
ABK99965  
ID ABK99965 standard; DNA; 2155 BP.  
XX  
AC ABK99965;  
XX  
DT 21-OCT-2002 (first entry)  
XX  
DE DNA encoding human secreted protein SCEP-48.

XX Secreted protein; SCEP; human; cell proliferative disorder; cancer;  
KW keratosis; arteriosclerosis; atherosclerosis; cirrhosis; hepatitis;  
KW psoriasis; autoimmune disorder; inflammatory disorder; AIDS; arthritis;  
KW acquired immunodeficiency syndrome; adult respiratory distress syndrome;  
KW Addison's disease; allergy; asthma; osteoporosis; autoimmune thyroiditis;  
KW Crohn's disease; dermatitis; diabetes; Graves' disease; haemodialysis;  
KW glomerulonephritis; scleroderma; systemic lupus erythematosus; uveitis;  
KW systemic sclerosis; ulcerative colitis; infection; trauma; Pick disease;  
KW cardiovascular heart disease; congestive heart failure; angina; epilepsy;  
KW hypertensive heart disease; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; amyotrophic lateral sclerosis; stroke; dementia;  
KW Huntington's disease; multiple sclerosis; neuromuscular disorder;  
KW metabolic disorder; endocrine disorder; toxic myopathy; mental disorder;  
KW schizophrenic disorder; developmental disorder; anaemia; epilepsy;  
KW hypothyroidism; glaucoma; sensorineural hearing loss; cataract;  
XX transgenic animal; gene; ds.

OS Homo sapiens.  
XX  
PN WO200248337-A2.  
XX

PD 20-JUN-2002.

XX 12-DEC-2001; 2001WO-US048517.

XX 13-DEC-2000; 2000US-0255639P.

PR 21-DEC-2000; 2000US-0257852P.

PR 05-JAN-2001; 2001US-0260105P.

PR 18-JAN-2001; 2001US-0282932P.

PR 18-JAN-2001; 2001US-0283096P.

PR 19-JAN-2001; 2001US-0263090P.

PR 02-FEB-2001; 2001US-0265926P.

XX (INCYTE GENOMICS INC.

XX Griffin JA, Yao MG, Duggan BM, Yue H, Ding L, Lal PG, Lee EA;

PI Ramkumar J, Thangavelu K, Xu Y, Lee S, Tang YT, Nguyen DB;

PI Warren BA, Honchell CD, Gietzen KJ, Baughn MR, Gandhi AR, Arvizu C;

PI Wallia NK, Lu Y, Elliott VW, Lu DAM, Hafalia AJA, Azimzai Y;

XX Khan FA, Tran UK;

XX WPI: 2002-583509/62.

DR P-PSDB; ABG69668.

XX Novel human secreted proteins and polynucleotides for diagnosing,

PT preventing or treating cell proliferative, autoimmune/inflammatory,

PT cardiovascular, neurological and developmental disorders.

XX Claim 5; Page 230; 234pp; English.

XX The invention describes an isolated polypeptide chosen from secreted

CC proteins (I), SEQ 1-54. (I) and the polynucleotide encoding it (II) are

CC useful for screening a compound for effectiveness as an agonist or

CC antagonist of (I) or compound that alters expression of (II). (I), the

CC identified agonist and antagonist are useful for treating a disease or

CC condition associated altered expression of functional SECP in a patient.

CC An antibody specific to (I) is useful for detecting the presence of (I),

CC purifying (I) from a sample and for diagnosing a condition or disease

CC associated with expression of SECP in a subject or in a biological

CC sample. (I) and (II) and modulators of (I) are useful for diagnosis,

CC treatment and prevention of cell proliferative disorders (e.g. cancer,

CC keratosis, arteriosclerosis, atherosclerosis, cirrhosis, hepatitis and

CC psoriasis), autoimmune/inflammatory disorders (e.g. acquired

CC immunodeficiency syndrome (AIDS), adult respiratory distress syndrome,

CC Addison's disease, allergies, asthma, osteoporosis, autoimmune

CC thyroiditis, Crohn's disease, dermatitis, diabetes, Graves' disease,

CC glomerulonephritis, arthritis, scleroderma, systemic lupus erythematosus,

CC systemic sclerosis, ulcerative colitis, haemodialysis, uveitis; viral,

CC bacterial, fungal, parasitic, protozoal, helminthic infections and

CC trauma), cardiovascular disorders (e.g. congestive heart failure, angina,

CC hypertensive heart disease), neurological disorders (e.g. Alzheimer's and

CC Pick disease, Parkinson disease, amyotrophic lateral sclerosis, epilepsy,

CC stroke, Huntington's disease, multiple sclerosis, dementia, neuromuscular

CC disorders, metabolic, endocrine and toxic myopathies, mental disorders,

CC schizophrenic disorders, and developmental disorders (e.g. anaemia,

CC epilepsy, hypothyroidism, glaucoma, sensorineural hearing loss and

CC cataract). (II) is useful for creating transgenic animals to model human

CC disease and to detect and quantify gene expression in biopsied tissues in

CC which expression of SECP is correlated with disease. This sequence

CC encodes a human secreted protein (SECP)

XX SQ Sequence 2155 BP; 488 A; 530 C; 545 G; 592 T; 0 U; 0 Other;

Query Match 49.5%; Score 190.6; DB 6; Length 2155;

Best Local Similarity 83.1%; Pseq No. 7.6e-39;

Matches 217; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 125 GCCTGCGACCAAGTTTCACAGAGGGTCTGAGGAATGAAGAGCCGTGGAAGGGCCACA 184

Db 205 GCCTGCGCTCCAGTTTCACAGAGGGTCTGAGGAATGAAGAGCCGTGGAAGGGCCACA 264

Qy 185 GCCATGTTGGTGTGAACGTGACAGTGGCCCTGTGGAGTGGAGGAGGGCCCGAG 244

Db 265 GCCACATGCAATGTAGCTGAGCAGAGGAGCCCTGTGGAGTGGAGGAAGGCTTGAG 324

QY 245 AACCTCAGAGATGGGACAGATACATCTCTGAGGCGAGGAGGGACACAGTGTGAGCTGCAG 304

Db 325 GCTCTCAGAGATGGGACAAATACAGCCTGAGACAGACGGGCTGTGTGTGAGCTGCAG 384

QY 305 ATCTGTGGCCTGCCATCGCGGAGCCCGGGAGTACTTGTGTGTGCGGCGAGGAGAGG 364

Db 385 ATCATGGCCTGGCTATGCGAGATACCGGGGTGACTCATGTGTGTGCGGCGAGGAGG 444

QY 365 ACCTCAGCAGCTCACCATC 385

Db 445 ACCTCAGCTACACTCACTGTC 465

# RESULT 10

ABN21414

ID ABN21414 standard; cDNA; 1005 BP.

XX AC ABN21414;

XX 24-JUN-2002 (first entry)

XX Human ORFX polynucleotide sequence SEQ ID NO:11305.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;

KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;

KW degenerative disorder; osteoarthritis; neurodegenerative disorder;

KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;

KW hypertension; hypothyroidism; cholesterol ester storage disease;

KW immune deficiency; immune disorder; infectious disease;

KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;

KW myasthenia gravis; gene; ss.

OS Homo sapiens.

XX WO200192523-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US010836.

XX 30-MAY-2000; 2000US-0206132P.

XX 29-AUG-2000; 2000US-0228716P.

XX (CURA-) CURAGEN CORP.

XX Shinkets RA, Leach MD;

XX WPI: 2002-106308/14.

XX P-PSDB; ABP05662.

XX Novel human polypeptides and polynucleotides useful for diagnosing,

PT preventing and treating cardiovascular disease, neurodegenerative,

PT hyperproliferative disorders and autoimmune disorders.

XX Disclosure; SEQ ID NO 11305; 1037pp; English.

XX The present invention describes substantially purified human proteins

CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1

CC in the specification). ABN15762 to ABN27252 encode the human ORFX

CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for

CC treating or preventing a pathology associated with an ORFX-associated

CC disorder in humans, and in the manufacture of a medicament for treating a

CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide

CC sequences can be used in gene therapy. ORFX sequences can be used in the

CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,

CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,

CC osteoarthritis, neurodegenerative disorders, disorders related to organ

CC transplantation, cardiovascular diseases, diabetes mellitus, systemic

CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester

CC storage disease, various immune deficiencies and disorders, infectious

CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid

CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host







CC with a disease state; and (3) identifying polynucleotides and  
CC polypeptides that predict compound sensitivity or resistance of cells  
CC associated with a disease state, comprising subjecting the plurality of  
CC cell lines to one or more compounds, analysing the expression pattern of  
CC a microarray of polynucleotides or polypeptides, and selecting  
CC polynucleotides or polypeptides that predict the sensitivity or  
CC resistance of cells associated with a disease state by using the  
CC expression pattern of the microarray. The polynucleotides and  
CC polypeptides have cytostatic activities, and can be used in gene therapy.  
CC The polynucleotides and polypeptides are useful in predicting the  
CC activity of compounds that interact with protein tyrosine kinases and/or  
CC protein tyrosine kinase pathways. These may be used in determining drug  
CC sensitivity in patients to allow the development of individualized  
CC genetic profiles which aid in treating diseases and disorders (e.g.  
CC cancer) based on patient response at a molecular level. The present  
CC sequence is used in the exemplification of the present invention.  
XX  
SQ Sequence 5382 BP; 943 A; 1699 C; 1798 G; 942 T; 0 U; 0 Other;

Query Match	12.4%;	Score 47.6;	DB 9;	Length 5382;
Best Local Similarity	49.0%;	Pred. No. 0.038;		
Matches 190;	Conservative 0;	Mismatches 189;	Indels 9;	Gaps 2;
QY	7	CTGTGAGCGAGGACGGGGCTGTGTGCGAGCTGCGAGATCTGTGGCTGGCTGTGGCAGATGC	66	
Db	2344	CTTGCAGAGGAAGACACATCTCGCGCCCTGGTGTGCGCGTGTCCAGTCTCAGACTC	2403	
QY	67	TGGGGAGTACTCCTGTGTGTGTGGGAGGAGAGA	126	
Db	2404	CGCGAGTACTTGTGTGAATTCACGATGATCGGCTCCTTCACTGTCACTGTCAAG	2463	
QY	127	CTGTCCAGCCAAATTACACAGAGGCTCT-----GAGGAATGAAGAGGCCGTGGAAGGGC	180	
Db	2464	ACCCGAGTGGGATCATATACCCCTCGGATGAGGTGACCTTGATGCCGTGACCTTGG	2523	
QY	181	CACAGCCATGTTGTGTGTAACCTGA---GCAAGGTGGCCCTGTGTGAGTGTGAGGAGGG	237	
Db	2524	GTGTGTGGTCTGATGTGTGAATCTGTCTCGGAGGATGCCCTGTGCGCTGGTACAAG	2583	
QY	238	GCCCGAGAACCTCAGAGATGGGACAGATACATCTTGAGCAGAGGGGACCAAGTGTGA	297	
Db	2584	TGGGCTGGAAATGGAGGAGACGAGGCCCTTGGTGTGTGAGAGGATGGGCCACGCTGCC	2643	
QY	298	GCTGTGAGATCTGTGGCTGCGCCATGGCGGACGCCGGGGAGTACTTGTGTGTGTGCGGGCA	357	
Db	2644	CTTGTGTCTACTTGTCTCTCAGCCGAGACGGGGGCGAGTTGTATGTGATCTGTGAGA	2703	
QY	358	GGAGAGAACCTCAGCCACGCTCAACATC	385	
Db	2704	TGACTCGGCGCTTCTTCACTGTCACTGTCT	2731	

RESULT 14	
AAS05390	
ID	AAS05390 standard; DNA; 81940 BP.
XX	
XX	AAS05390;
XX	
XX	24-OCT-2001 (first entry)
XX	
XX	Human titin (connectin) gene sequence.
XX	
XX	Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;
KW	titin-related disease; zebrafish; heart failure; heart disease; ds.
KW	
XX	
XX	Homo sapiens.
OS	
XX	
XX	Key
XX	Location/Qualifiers
PH	133..80913
CDS	/*tag= a
FT	/product= "titin"
FT	
FT	
XX	
XX	
PN	WO200151666-A1.

XX 19-JUL-2001.  
XX PD  
XX 12-JAN-2001; 2001WO-US001212.  
XX PF  
XX 12-JAN-2000; 2000US-0175787P.  
XX PR  
XX (GEHO ) GEN HOSPITAL CORP.  
XX PA  
XX Fishman MC;  
XX PI  
XX WPI; 2001-451869/48.  
XX DR  
XX P-PSDB; AAU05396.  
XX DR  
XX Determining if a subject has or is at risk of developing a titin-related  
XX disease or condition, particularly heart failures, comprises detecting  
XX the presence of a mutation in the titin gene.  
XX PT  
XX PS Disclosure: Page 35-57; 114pp; English.

XX The present sequence encoding for human titin (also known as connectin)  
CC is described in an invention relating to a novel method for determining  
CC whether a subject has or is at risk of developing a titin-related disease  
CC or condition. The method comprises analysing a nucleic acid of a sample  
CC from the subject and detecting the presence of a mutation (e.g. the  
CC pickwick mutation in the cardiac specific exon N2B) in the titin gene,  
CC which indicates that the subject has or is at risk of developing a titin-  
CC related disease. The zebrafish which has a phenotype similar to mammalian  
CC heart failure is used as a model. The method is useful for detecting an  
CC increased likelihood of heart disease, such as heart failure, in a  
CC patient, so that appropriate intervention can be instituted before any  
CC symptoms occur. The method may also be used to facilitate determination  
CC of etiology of an existing heart condition, such as heart failure, to  
CC identify compounds that can be used to treat or prevent heart conditions,  
CC in prenatal genetic screening, e.g. to identify parents who may be  
CC carriers of a recessive titin mutation. Compounds identified using the  
CC methods may be used to treat patients that have or are at risk of  
CC developing heart disease, e.g. heart failure  
XX  
XX Sequence 81940 BP: 263373 A: 17100 C: 19899 G: 19568 T: 0 U: 0 Other:

Query Match	12.2%;	Score	46.8;	DB	4;	Length	81940;
Best Local Similarity	48.8%;	Pred.	No. 0.1;				
Matches	157;	Conservative	0;	Mismatches	162;	Indels	3;
Gaps	1;						

  

Qy	66	CTGGGGAGTACTCTGTGTGTGTGGGGAGGAGAGACCTCTGCCACTCTCACCGTGAAGG	125
Db	16091	CAGGAGAGGTTTCCTTCACGGCTGTATGCCAAATCTGCAGCCAACTCGAAGTGAAG	16150
Qy	126	CCCTGCCACCAGAGTTTCACAGAGGCTCTGAGGAATGAAGAGCCGTGGAGGGGCCACAG	185
Db	16151	AATTGCCCTCTTATCTTCATCACACCTCTCAGTGATGTTAAAGTCTTCGAGAAAGATGAGG	16210
Qy	186	CCATGTTGTGTGTGTAACTGACCAAGGTGGCC---CCTGTGGAGTGGAGGAAGGGGCCG	242
Db	16211	CTAAGTTTCAGTGTGAAGTATCCAGGAGGCCAAAACATTCCTGTGGCTAAAGGAACCC	16270
Qy	243	AGAACCTCAGAGATGGGNCAGATACATCTCTGAGCAGGAGGGACAGGTGTGAGCTGC	302
Db	16271	AGAAATCAGAGTATGATGACATTTGACCTTATTAAGGATGGCACTAAGCATTCATGG	16330
Qy	303	AGATCTGTGGCTTGGCCATGGCGGACGCCGGGGAGTACTTGTGTGTGCGGGCAGGAGA	362
Db	16331	TGATCAAGTCAGCTGCTTTTGAAGATGAAGCAAAATACATCTTTGAAGCTGAAGATAAGC	16390
Qy	363	GGACCTCAGCAGGCTCACCAT	384
Db	16391	ACACAAAGTGCAACTGATCAT	16412

  

RESULT	15
ABK64829	
ID	ABK64829 standard; DNA; 81940 BP.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 12:53:28 ; Search time 441.18 Seconds  
(without alignments)  
10303.209 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_1\_1070

Perfect score: 1070

Sequence: 1 tgcctaccagcagccacac.....tagtgcgagccgcgggt 1070

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1070	100.0	24120	7 ABX11642	Abx11642 Human ser
2	255	23.8	2884	7 ADA53433	Ada53433 Human cod
3	197.6	18.5	9591	4 AAS28906	Aas28906 Human imm
4	197.6	18.5	9591	4 AAL03355	Aal03355 Human rep
5	197.6	18.5	9591	9 ADB31747	Adb31747 Human nov
6	197.6	18.5	12415	4 AAS28907	Aas28907 Human imm
7	197.6	18.5	12415	4 AAL03356	Aal03356 Human rep
8	197.6	18.5	12415	9 ADB31748	Adb31748 Human nov
9	197.6	18.5	20565	4 AAK89083	Aak89083 Human dig
10	197.6	18.5	20565	4 AAS28908	Aas28908 Human imm
11	197.6	18.5	20565	4 AAL03357	Aal03357 Human rep
12	197.6	18.5	20565	9 ADB31749	Adb31749 Human nov
13	126	11.8	5382	9 ADD14722	Add14722 Human src
14	117.2	11.0	471	2 AAV88238	Aav88238 EST clone
15	99.6	9.3	9975	7 AAL61173	Aal61173 Actinosyn
16	99.6	9.3	82746	7 AAL61174	Aal61174 Actinosyn
17	90.2	8.4	114955	2 AAX53491	Aax53491 Human ade
18	86.6	8.1	609	4 ABA06472	Aba06472 Human cod
19	86.6	8.1	609	6 ABV83809	Abv83809 Human pol
20	83.2	7.8	114955	2 AAX53491	Aax53491 Human ade
21	78.6	7.3	14055	7 AAL61170	Aal61170 Actinosyn
22	74.8	7.0	3015	3 ABA29367	Aba29367 Human zsi
23	73.4	6.9	65140	4 AAD17184	Aad17184 Streptomy

24	73.4	6.9	125401	4	AAD17186	Aad17186 Streptomy
25	72.8	6.8	3183	7	ABZ77314	Abz77314 Nucleotid
26	70.6	6.6	4020	2	AAT91361	Aat91361 Orf virus
27	70.2	6.6	1281	7	ABZ66730	Abz66730 Orthosomy
28	70.2	6.6	15240	7	ABZ66809	Abz66809 Orthosomy
29	70.2	6.6	109519	5	AAS08693	Aas08693 Micromono
30	70	6.5	4341	7	ABT32151	Abt32151 Benzodiaz
31	70	6.5	32539	7	ABT32129	Abt32129 Benzodiaz
32	68.6	6.4	1614	5	AAB26499	Aab26499 Human low
33	68.6	6.4	5760	6	ABST78661	Abst78661 M. echino
34	68.6	6.4	12425	5	AAB26495	Aab26495 Human low
35	68	6.4	1226	4	AAL07456	Aal07456 Human rep
36	68	6.4	1226	4	AAL07456	Aal07456 Human rep
37	68	6.4	1776	9	ADE03416	Ade03416 Human imm
38	68	6.4	1779	6	AAD16344	Aad16344 Human sbg
39	68	6.4	1811	6	ABQ86161	Abq86161 Novel hum
40	68	6.4	1811	6	ABK48387	Abk48387 DNA encod
41	68	6.4	1860	6	ABK92037	Abk92037 DNA encod
42	68	6.4	2026	9	ADB62562	Adb62562 Human cdn
43	68	6.4	2080	6	ABK62080	Abk62080 Human cdn
44	68	6.4	2282	6	ABK62081	Abk62081 Human cdn
45	68	6.4	2300	6	AAD41275	Aad41275 Human LP
						Abq77432 Human CGD

## ALIGNMENTS

## RESULT 1

ABX11642  
ID ABX11642 standard; cDNA; 24120 BP.

AC ABX11642;

DT 09-MAY-2003 (first entry)

DE Human serine/threonine or protein kinase 12599, cDNA.

Human; ss; gene; serine/threonine kinase; protein kinase; 12599;  
cardiovascular disease; heart failure; myocardial infarction;  
blood vessel disorder; atherosclerosis; Kaposi's sarcoma;  
hemolytic anaemia; cellular proliferative disorder; cancer;  
protein kinase disorder; autoimmune disorder; diabetes mellitus;  
psoriasis; inflammatory bowel disease; rheumatoid arthritis;  
multiple sclerosis.

OS Homo sapiens.

Key	Location/Qualifiers
5'UTR	1..71
CDS	/*tag= a 72..23978 /*tag= b /product= "Kinase 12599" /note= "This CDS is specifically claimed in claim 2" 23979..24120 /*tag= c
3'UTR	
US2002168742-A1.	
14-NOV-2002.	
15-FEB-2002; 2002US-00077130.	
15-FEB-2001; 2001US-0269201P.	
(MILL-) MILLENNIUM PHARM INC.	
Kapeller-Libermann R, Acton SL;	
WPI: 2003-298729/29.	
P-PSDB; ABG76187.	

PT Novel isolated human protein kinase, designated 59079 or 12599  
 PT polypeptide, useful as diagnostic and therapeutic agents for preventing  
 PT cardiovascular diseases, proliferative disorders, and protein kinase  
 PT disorders.

PS Claim 2; Page 58-84; 119pp; English.

XX The invention relates to an isolated human serine/threonine or protein  
 CC kinase, 59079 or 12599 polypeptide, encoded by nucleic acid molecule  
 CC comprising at least 85% identity to the nucleic acids appearing as  
 CC ABX11641 and ABX11642 or their complement, a naturally occurring variant  
 CC of the kinases or their fragments. Also included are a non-human host  
 CC cell containing the nucleic acids, an antibody specific for the proteins,  
 CC identifying a compound which binds to the kinase (by contacting the  
 CC kinase or a cell expressing the kinase with a test compound and  
 CC determining whether the kinase binds to the test compound) and modulating  
 CC the activity of kinase using the identified compound. The kinases and  
 CC their encoding nucleic acids are useful as diagnostic and therapeutic  
 CC agents for preventing a disease or condition associated with an aberrant  
 CC or unwanted 59079 or 12599 activity in a subject, including  
 CC cardiovascular diseases such as heart failure, and myocardial infarction;  
 CC disorders involving blood vessels such as atherosclerosis, and Kaposi's  
 CC sarcoma; blood platelets disorder such as thrombocytopenia, leukaemia,  
 CC Hodgkin's disease, haemolytic anaemia; cellular proliferative disorders  
 CC such as cancer; and protein kinase disorders such as autoimmune  
 CC disorders, diabetes mellitus, psoriasis, inflammatory bowel disease,  
 CC rheumatoid arthritis, and multiple sclerosis (many examples of diseases  
 CC and disorders are included in the specification). The kinases, their  
 CC encoding nucleic acids and antibodies are useful in screening assays,  
 CC detection assays (e.g. forensic biology), and predictive medicine (e.g.  
 CC diagnostic assays, prognostic assays, and monitoring clinical trials and  
 CC pharmacogenomics). The kinases and their encoding nucleic acids are  
 CC useful as query sequences to perform a search against public databases to  
 CC identify other family members or related sequences. The present sequence  
 CC encodes the kinase 12599

SQ Sequence 24120 BP; 4719 A; 7265 C; 8278 G; 3858 T; 0 U; 0 Other;

Query Match 100.0%; Score 1070; DB 7; Length 24120;  
 Best Local Similarity 100.0%; Pred. No. 6.8e-162;  
 Matches 1070; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCCTACAGCAGCCACACTCCCGCGCTGCCAGAGCCCCCATAGAGAGGTCCCCG 60  
 DB 1 TGCCTACAGCAGCCACACTCCCGCGCTGCCAGAGCCCCCATAGAGAGGTCCCCG 60  
 QY 61 CCGCACCGTATGATCAGGCACAGTTCAGCGGGGGCGCCGCTTCTCACCGGCCCA 120  
 DB 61 CCGCACCGTATGATCAGGCACAGTTCAGCGGGGGCGCCGCTTCTCACCGGCCCA 120  
 QY 121 AGGCCTTCGTGGTGGGCAAGGACGACACCTTCAGTGCAGATCGTGGTAATC 180  
 DB 121 AGGCCTTCGTGGTGGGCAAGGACGACACCTTCAGTGCAGATCGTGGTAATC 180  
 QY 181 CCAGCCACAGGTGAGTGGGAGAGGACGACGAGCGGTGACGCGCGCGCGCTTCC 240  
 DB 181 CCAGCCACAGGTGAGTGGGAGAGGACGACGAGCGGTGACGCGCGCGCGCTTCC 240  
 QY 241 GTCTGGCCACGAGCGGCGACCTTACCGCTTACCTTGGACCTGGCGCTGGCGACA 300  
 DB 241 GTCTGGCCACGAGCGGCGACCTTACCGCTTACCTTGGACCTGGCGCTGGCGACA 300  
 QY 301 GTGGGCAATACGTGTGCGCGCGGCAATGCCATAGCGAGGCTTCGCTGCCGTGGCC 360  
 DB 301 GTGGGCAATACGTGTGCGCGCGGCAATGCCATAGCGAGGCTTCGCTGCCGTGGCC 360  
 QY 361 TGCAGGTGAGCGGAGCGCGGTGGCGGAGGAGGACGACCTTCCTGCTGCGGCCCA 420  
 DB 361 TGCAGGTGAGCGGAGCGCGGTGGCGGAGGAGGACGACCTTCCTGCTGCGGCCCA 420  
 QY 421 CGTCCATCCGCGTGGCGAGGCTCAGAGGCCACTTTCGCTGCCGTGGGTCC 480  
 DB 421 CGTCCATCCGCGTGGCGAGGCTCAGAGGCCACTTTCGCTGCCGTGGGTCC 480

QY 481 CGAGCCCGCAGTGAAGTGGTCCAAAGACCGGCGCGCTGGGTGAAGCCGAGCGGCC 540  
 DB 481 CGAGCCCGCAGTGAAGTGGTCCAAAGACCGGCGCGCTGGGTGAAGCCGAGCGGCC 540  
 QY 541 CGCTCGCGTGGAGAGCTCGCGGAGGAGTGGCTCGCATTCGGCGCGCGCGCGCG 600  
 DB 541 CGCTCGCGTGGAGAGCTCGCGGAGGAGTGGCTCGCATTCGGCGCGCGCGCGCG 600  
 QY 601 CGAGCGCGGACATTTACGAGTCCGCGCGGAGAACCCGCTGGGCGCTGCCAGCCCG 660  
 DB 601 CGAGCGCGGACATTTACGAGTCCGCGCGGAGAACCCGCTGGGCGCTGCCAGCCCG 660  
 QY 661 CGGCGTGTGTGAGTTCGAGACCGCGCGGAGACGCGGCGCGCGCGCGGACCTCCAG 720  
 DB 661 CGGCGTGTGTGAGTTCGAGACCGCGCGGAGACGCGGCGCGCGCGGACCTCCAG 720  
 QY 721 CGGCGTCTCTGGCGCACCTGCAGCGCGCGCGGAGGCTATCGCGCGCGCGCGCG 780  
 DB 721 CGGCGTCTCTGGCGCACCTGCAGCGCGCGCGGAGGCTATCGCGCGCGCGCGCG 780  
 QY 781 CCTACCCCGCAGCACCGGACCGCGCGACCTGCAGCGGTGACTGAAGGCAAGCACCGCG 840  
 DB 781 CCTACCCCGCAGCACCGGACCGCGCGACCTGCAGCGGTGACTGAAGGCAAGCACCGCG 840  
 QY 841 TCAGCTGTGTGAGTCCGCGCGGAGCCCAAGCCGAGAGCGGTGTGGAAGAGGAGCGCG 900  
 DB 841 TCAGCTGTGTGAGTCCGCGCGGAGCCCAAGCCGAGAGCGGTGTGGAAGAGGAGCGCG 900  
 QY 901 TGGTCAACGAGGCGCGCGCCCACTGTGTGTACGAGGACGCGCAGGAACTTCGTGCTCA 960  
 DB 901 TGGTCAACGAGGCGCGCGCCCACTGTGTGTACGAGGACGCGCAGGAACTTCGTGCTCA 960  
 QY 961 AGATCTCTTCTGCAAGCAGTTCGGACCGCGCGCTTACCTGACGCGGCTCCAACTCG 1020  
 DB 961 AGATCTCTTCTGCAAGCAGTTCGGACCGCGCGCTTACCTGACGCGGCTCCAACTCG 1020  
 QY 1021 TGGGCGACACCTACAGCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1070  
 DB 1021 TGGGCGACACCTACAGCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1070

## RESULT 2

ADA53433/c

ID ADA53433 standard; cDNA; 2884 BP.

XX ADA53433;

XX AC

XX DT 20-NOV-2003 (first entry)

XX DE Human coding sequence, SEQ ID 1001.

XX KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;

XX KW Gene Therapy; human; secretory protein; membrane proteins; cancer;

XX KW inflammatory disease; osteoporosis; neurological disease; gene; ss.

XX OS Homo sapiens.

XX PN BP1293569-A2.

XX PD 19-MAR-2003.

XX PF 21-MAR-2002; 2002EP-00006586.

XX PR 14-SEP-2001; 2001JP-00328381.

XX PR 24-JAN-2002; 2002US-0350435P.

XX PA (HELI-) HELIX RES INST.

XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

XX WPI; 2003-395539/38.  
DR P-PSDB; ADA55072.  
XX  
PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in  
PT which the gene is involved, or as target molecules for gene therapy.  
XX  
PS Claim 1; SEQ ID NO 1001; 205pp; English.  
XX  
CC The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
SQ Sequence 2884 BP; 629 A; 798 C; 779 G; 678 T; 0 U; 0 Other;  
Query Match 23.8%; Score 255; DB 7; Length 2884;  
Best Local Similarity 100.0%; Pred. No. 3.9e-32;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 805 GCACCTGACGGTGAAGCAAGCAGCGCGCCTCAGTGTCTAGTACCGGCGAGC 864  
Dd |||||  
634 GCACCTGACGGTGAAGCAAGCAGCGCGCCTCAGTGTCTAGTACCGGCGAGC 575  
QY 865 CCAAGCCGACGCGTGTGAAGAGGACGCGCAGCTGTGACCGAGGCGCGCCACG 924  
Dd |||||  
574 CCAAGCCGACGCGTGTGAAGAGGACGCGCAGCTGTGACCGAGGCGCGCCACG 515  
QY 925 TGGTGTACGAGGACGCGCAGGAGAACTTCTGCTCAAGATCCTTTCTGCAAGCAGTCGG 984  
Dd |||||  
514 TGGTGTACGAGGACGCGCAGGAGAACTTCTGCTCAAGATCCTTTCTGCAAGCAGTCGG 455  
QY 985 ACCGGCGCCTTACACCTGCACGCGCTCCACCTCTGCGGCCAGACCTAGCTCTGTGC 1044  
Dd |||||  
454 ACCGGCGCCTTACACCTGCACGCGCTCCACCTCTGCGGCCAGACCTAGCTCTGTGC 395  
QY 1045 TGGTCGTAGTCGCG 1059  
Dd |||||  
394 TGGTCGTAGTCGCG 380  
RESULT 3  
AAS28906  
ID AAS28906 standard; DNA; 9591 BP.  
XX  
AC AAS28906;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immunoglobulin encoding genomic DNA SEQ ID No 268.  
XX  
KW Immunoglobulin; signal transduction pathway protein; cancer; ds;  
KW antisense therapy; gene therapy; neurological disorder; renal disorder;  
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;  
KW reproductive disorder; immune system disorder; proliferative disorder;  
KW muscular disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200155315-A2.  
XX  
FD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001326.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 27-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216800P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 12-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.

PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241212P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249298P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 01-DEC-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254037P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-457725/49.  
XX  
XX Isolated novel immunoglobulin polypeptide for monitoring the presence and  
XX progression of diseases and for diagnosis.  
XX  
XX Claim 1; SEQ ID NO 268; 551pp; English.  
XX  
XX Sequences AAS28878-AAS28926 represent genomic DNA molecules which encode  
XX the immunoglobulin polypeptides of the invention. The polynucleotides and  
XX polypeptides can be used to diagnose a pathological condition or a  
XX susceptibility to a pathological condition in a subject by determining

CC the presence or absence of a mutation in a DNA sequence or determining  
CC the presence or amount of expression of the protein. Alternatively the  
CC identification of a binding partner to a sequence allows determination of  
CC changes in protein activity. The sequences can be used as research tools  
CC for receptors or other signal transduction pathway proteins that interact  
CC with the polypeptides of the invention and can be used to treat, prevent  
CC or diagnose various types of disorders such as neurological disorders,  
CC cardiovascular disorders, gastrointestinal disorders, reproductive  
CC disorders, immune system disorders, renal disorders, muscular disorders,  
CC pulmonary disorders, proliferative disorders and cancer. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 9591 BP; 2026 A; 2759 C; 2671 G; 2135 T; 0 U; 0 Other;

Query Match 18.5%; Score 197.6; DB 4; Length 9591;  
Best Local Similarity 51.7%; Pred. No. 4.8e-23;  
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

QY 7 CCAGCAGCCACACTCCGGCGCTGCCAGAGCCCTCATAGAGAGAGTCCCGCCGCCA 66  
DB 2 CCTTCGGCACCCCTCCCTGCCCCCAGCGTTCCGCGCTGCAGCGGTCCGCGCGCGA 61  
QY 67 CGGTCAAGGATACGCCACAGTTCAGCGGGGGCCCCCTTTCTCACCGGGCCCAAGCCT 126  
DB 62 TGAAGCGAGCTCGGGGATCAGGGGAGCCCCCGTCTCTCGCTTCCCGGGCGCTG 121  
QY 127 TCGTGTCTCGGTGGCAAGGACGCCACCTCAGCTGCCAGATCGTGGTAATCCACGC 186  
DB 122 TCGGGGTGTAAAGTGGCCCGGAGCCGAGCTCAAGTGTGTGTCTCTGGGAGCGCGCC 181  
QY 187 CACAGTCAAGTGGGAGAGAACAGCAGCGCGGTGACGGCGGGCGCGCTTCGTCTGG 246  
DB 182 CTGTAGTGTGTGGGAGAGGGGGGAGCAGCTGGCGGCTCGGAACGCTGAGCTTC 241  
QY 247 CCAGGACGGGACCTTACCGCTCCTACTATCTTGGACCTGGCGCTGGGCGACAGTGGC 306  
DB 242 CGCGGACGGCGGAGCAGCGCTGTGTGACCGCGGCACTCCACCGACGCGGGGG 301  
QY 307 AATACGTGTGCGCGCGCGCAATGCCATAGCGAGGCTTCGTCTGCGTGGCGCTGCAGG 366  
DB 302 TCTAGTGTGTGCGCGCGCGCAACGCGCGCGAGGCTAGCGCGCGCGCGCTCACCG 361  
QY 367 TGGACGCGAGGCGCGGTGCGCGAGAGGCGCGCGACTTCTCTGCTGGGCGCCACGTCCA 426  
DB 362 TGCTGGAGCGCGCGCTCCGACCCCGAGCTGACAGCCCGCGAGCGCGCTGCATCGC 421  
QY 427 TCCGCGTGGGAGGCTCAGAGGCCACCTTCGCTGCGG-----CGTGGGTGGTCCC 480  
DB 422 CGGGTCCGGGAGGGGCGCGGCTTCTCTACGGGGCTCGATCCAGTGGTGTGTC 481  
QY 481 CGAGCCCGCAAGTGTGTTCAGGAGCGCGCGCGCTGGGTGAGCCGACGCGGCCCC 540  
DB 482 GGGGGCGGAGTGTGTGACGTGCGGGCGGGGGGCTCCCGAGCCACACATGTACT 541  
QY 541 CGGTGCGGTGAGGAGCTCGCGAGGCAAGTGTGCGCTGCGCATTCGGCGCGCGCGCG 600  
DB 542 GGGAGAGGACGGATGGCCCTTGACGAAGTGTGGACAGAGCCATTCGCGTCCAGC 601  
QY 601 GCGAGCGCGCACTTACAGGTCCCGCGCGAGAACCCGCTGGG--GCTGCGACGCGCGC 658  
DB 602 CGGCGCGCGAGGAGCGCGCGCGGAGCTGGCACTGCGCATCTCTGCGGCTCGGC 661  
QY 659 GCGCGCGTGTG--GTGACTCGAGACCGCGGACAGGCGCGCGCGCGCGCGCGCTCCA 717  
DB 662 TGC CGGATTCGCGGCTGTACGTGTGCACGCGCGCGCGCGCGCGCGCGCGCGCGG 721  
QY 718 CGGCGCGCTCTCTGCGCACCTTGACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 777  
DB 722 GGGCGCTCTCCAGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 781  
QY 778 CGGCTC-----ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 828

Db 782 CCGCGCGGTGGTGGAGCCGCTCAAGTGCAGCGCTTAAGACCTTCTGGGTGAACGAGGGCA 841  
Qy 829 AGCACGCGCGCTCAGCTGTCTAGTACCGCGCGAGCCCAAGCCGAGACCGGTGTGGAAGA 888  
Db 842 AGCAGCCAAAGTTCGCTGCTACGTGATGGCAAGCCGAGCCGAGATCGAATGCACT 901  
Qy 889 AGGACGCGCAGCTGGTACCGAGGCGCGCGCCACAGTGTGTAGAGGAGCGCGCAGAGA 948  
Db 902 GGGAGGCGCGCGCGCTCTCCGAGCCGCGCGCCCTCATGTACCGCGACCGCGAGCGCGG 961  
Qy 949 ACTTCGTGCTCAAGATCCTCTCTGCAAGCAGTCCGACCGCGGCTCTACCTGTGACACGG 1008  
Db 962 GCTTCGTGCTCAAGTCTTACTGCGAGCCAGGATCGTGGCTCTAGCTGTGGCGG 1021  
Qy 1009 GGTCCAACTGTGGGCGACACCTACAGCTCTGTGCTGTGAGTGGCGGAGCCCGCGG 1068  
Db 1022 CGCGAACTCGCGCGCGCAGACGCTCAGTGCCTGCACTGCAGTGAAGGTACGCGCGG 1081

RESULT 4  
AAL03355  
ID AAL03355 standard; DNA; 9591 BP.  
XX AC AAL03355;  
XX DT 21-NOV-2001. (first entry)  
XX DE Human reproductive system related antigen DNA SEQ ID NO: 6043.  
XX KW Human; reproductive system related antigen; reproductive system disorder;  
XX KW cancer; gene therapy; ds.  
XX OS Homo sapiens.  
XX PN WO20015320-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US001339.  
XX PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0194664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246603P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 17-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.

PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 05-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254087P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-465570/50.  
 XX  
 PT Isolated nucleic acid molecule encoding a reproductive system antigen is  
 PT used in preventing, treating or ameliorating a medical condition.  
 PT  
 PS Disclosure; SEQ ID NO 6043; 1297pp + Sequence Listing; English.  
 PS  
 XX The present invention provides the protein and coding sequences of a  
 CC number of human reproductive system related antigens. These can be used  
 CC in the prevention and treatment of reproductive system disorders,  
 CC including cancer. The present sequence is a genomic sequence encoding a  
 CC protein of the invention  
 XX  
 SQ Sequence 9591 BP; 2026 A; 2759 C; 2671 G; 2135 T; 0 U; 0 Other;  
 Query Match 18.5%; Score 197.6; DB 4; Length 9591;  
 Best Local Similarity 51.7%; Pred. No. 4.8e-23;  
 Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;  
 QY 7 CCACGAGCCACACTCCGGCGGTGCCCCAGAGCCCCCATAGAGAGGTCCCCCGCCGCA 66  
 DB 2 CCTCCGACCCCCCTGCCCCCCACCGTTGCGCGCTGCAGGCGGTGGCGCGCCGGA 61  
 QY 67 CGGTATGATGACCCACAGTTACAGCGGGCGCCCGCTTTCTACCCCGGCCCAAGCCT 126  
 DB 62 TGAAGGCGAGTCCGGGGGATCAGGGAGCCCCCGTCTTCTCGCGCTTCCCGCGCCTG 121  
 QY 127 TCGTGGTGTGGTGGGCAAGAGCCACCTCAGTGCACAGTCTGGGTAAATCCACGC 186  
 DB 122 TCGGGTGGTAAAGTGGCGCCGAGGCGAGTCAAGTGGTCTTGGGGAGCCCGCCG 181  
 QY 187 CACAGGTGAGTGGAGAGGACACAGCAGCGGTGACGCGCGCGCGCTCCGCTCG 246  
 DB 182 CTGTAGTGGTGGAGAGGGGGGAGCAGAGTGGCGGCTCGGAACGCTGAGCTTCC 241  
 QY 247 CCCAGGACGCGACCTCTACCGCTCACTCTGACCTGGCGCTGGGCGACAGTGGGC 306

Db 242 CGGCGACGCGCGGACACGGCCTGTGTGACCGCGCACTGCCACACGACGCGGGG 301  
 QY 307 AATAGTGTGCGCGCGGCAATGCCATAGCGAGGCTTGGCTGCGTGGGCTGAGG 366  
 Db 302 TCTAGTGTGCGCGCGGCAACGCGCGGAGGCTACGCGCGCGCGCGCTACCG 361  
 QY 367 TGGACGCGGAGCGCGCTGCGCGCGAGAGGCGCGCACTTCTGCTCGGCGCCACGTC 426  
 Db 362 TGCTGAGCCCGCGCTCCGACCCCGAGCTGCAGCCCGCGCGCGCTGCAATCG 421  
 QY 427 TCCGCGTGGGAGGCTCAGAGCCACTTCCGCTGCGG-----GTTGGTGGCTCC 480  
 Db 422 CGGGTCCGGGAGGCGCGCGCTTCTCTCAGGGGCTTCGATCCAGTGGGTGCTG 481  
 QY 481 CGAGCGCGGAGTGTGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 540  
 Db 482 GGGGGGAGGAGTGTGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 541  
 QY 541 GGTGCGGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 600  
 Db 542 GGGAGAGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 601  
 QY 601 GCGAGCGCGCACTTACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 658  
 Db 602 CGGCGCGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 661  
 QY 659 GCGCGCGCTAGTG-GTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 717  
 Db 662 TGCGGATTCGGCGTCTACGTGTGCGCACGCCCGCAACGCGCACGCGCAGCGG 721  
 QY 718 CGGCGCGCTCTGCGGCACTTACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 777  
 Db 722 GGGCGCTGTCCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 781  
 QY 778 CCGCTC-----ACCGCGCAGCACCGCGCACCTGTCACGCTGACTGAAGGCA 828  
 Db 782 CCGCGCGGTGTGAGCGCGCTCAAGTGTGCGCTTAAGACCTTCTGGTGAACGAGGCA 841  
 QY 829 AGCAGCGCGCTCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 888  
 Db 842 AGCAGCGCAAGTTCGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 901  
 QY 889 AGGAGCGCACTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 948  
 Db 902 GGGAGCGCGCGCTGCTCCGACCGCGCGCTCATGTACCGGACCGCGACGCGGCG 961  
 QY 949 ACTTGTGCTCAAGATCTTCTGCAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1008  
 Db 962 GTTCTGTCTCAAGTGTCTTACTGCGAGGCGCGAGGATCGTGGGCTTACGCTGTGCG 1021  
 QY 1009 CGTCCACCTGTGCGGCGAGCTTACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1068  
 Db 1022 CGGCACTCGCGCGGCGAGCTCAGTGTGCGGCTGACGCTGACGCTGACGCTGACGCG 1081  
 RESULT 5  
 ADB31747  
 ID ADB31747 standard; DNA; 9591 BP.  
 XX  
 AC ADB31747;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human novel protein DNA SEQ ID NO 268.  
 XX  
 DE gene therapy; human; immunoglobulin; cancer; lung cancer; leukaemia;  
 KW ovarian cancer; epithelial cancer; hyperplasia; Gaucher's disease; AIDS;  
 KW arhythmia; cardiac oedema; ischaemia; pneumonia; cystic fibrosis;  
 KW asthma; sarcoidosis; rhinitis; anaemia; inflammation; sinusitis;  
 KW chronic obstructive pulmonary disease; infectious disease; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX



[illegible]

PI Rosen CA, Ruben SM, Barash SC;  
XX WPI; 2003-615993/58.  
XX  
XX New human immunoglobulin superfamily of polypeptide and genes, useful for  
PT treating, preventing or diagnosing e.g. cancers (lung cancer, leukemia,  
PT ovarian epithelial cancer, etc.), hyperplasia, ischemia, pneumonia or  
PT AIDS.  
XX  
XX Disclosure; SEQ ID NO 268; 213pp; English.  
XX  
XX The invention relates to an isolated polypeptide, which comprises the  
CC human immunoglobulin superfamily of proteins. The polypeptide or  
CC polynucleotide is useful for treating, preventing or ameliorating a  
CC medical condition e.g. cancers (lung cancer, leukemia, AIDS,  
CC epithelial cancer, etc.), hyperplasia, Gaucher's disease, ovarian  
CC arhythmia, cardiac oedema, ischaemia, pneumonia, cystic fibrosis,  
CC asthma, sarcoidosis, rhinitis, anaemia, leukaemia, inflammatory  
CC sinusitis, chronic obstructive pulmonary disease, infectious diseases  
CC etc. The polypeptide or polynucleotide is also useful for diagnosing any  
CC of these diseases or a susceptibility to the disease. The present  
CC sequence represent a novel human protein DNA. Note: The sequence data for  
CC this patent did not form part of the printed specification but was  
CC obtained in electronic format direct from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030077606  
XX  
SQ Sequence 9591 BP; 2026 A; 2759 C; 2671 G; 2135 T; 0 U; 0 Other;  
Query Match 18.5%; Score 197.6; DB 9; Length 9591;  
Best Local Similarity 51.7%; Pred. No. 4.8e-23;  
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

QY 7 CCACAGCGCCACATCCGGCGGTGCCCCAGAGCCCCCATAGAGAGAGTCCCGCGGCCA 66  
Db 2 CCCTCCGACCCCCCTGCCCCCCCCACCGTTCCGCGCTGCAGGCGGTCCGCGCGCGCA 61  
QY 67 CCGTCATGTGATCACCACAGTTACAGCGGGCGCGCCCGCTTCTCACCGCGCCCAAGCCT 126  
Db 62 TGAAGGCGAGCTCGGGGATCAGGGAGCCCCCGTCTCTCGCGCTTCCCGCGCCTG 121  
QY 127 TCGTGTGTCGTGGGCAAGACGCCACCTCAGTCCAGATCGTGGGTAATCCACGC 186  
Db 122 TCGGGTGGTAAGTGGCGCGGAGCGCGAGTCAAGTGCCTGCTGGGGAGCGCGCGC 181  
QY 187 CACAGGTGAGTGGGAGAGGACACAGCAGCGGTGACGGCGCGCGCTTCCGCTGG 246  
Db 182 CTGTAGTGTGTTGGAGAGGGCGGGAGCAGTGGCGGCTCGGAACGCTGAGTTCC 241  
QY 247 CCCAGGACGGGACCTCTACCGGCTCTACTCTCTGACCTGGGCTGGGACAGTGGGC 306  
Db 242 CGGCGGACGGCGGAGCAGCGCTGTGTGACCGCGCGCACTGCGCACCGCGGGG 301  
QY 307 AATACGTGTCGCGCGCGCAATGCCATAGCGAGGCTTCGCTGCGTGGGCTCAGG 366  
Db 302 TCTACGTGTCGCGCGCGCAACCGCGCGCGGAGGCTTACGCGGCGCGCGCGCTCACCG 361  
QY 367 TGAACGCGGAGGCGCGTGGCGGAGCAGCGCGCGCACTTCTGCTGGCGCCACGCTCCA 426  
Db 362 TGCTGGAGCGCGCGCTCCGACCCGAGTGCAGCGCGCGCGAGCGCGCTGCCATCGC 421  
QY 427 TCCGCTGTCGCGAGGCTCAGAGCCACTTCCGCTGCGG- - - - -CGTGGGTGGCTCC 480  
Db 422 CGGGGTCCGGGAGGCGCGCGCGCTTCTCTCAAGGCGCGCTCGATCCAGTGGGTCTGC 481  
QY 481 CGAGCGCGGAGTGAAGTGTGCTTCAAGACGCGGCGCGCGCTGGGTGAGCGCGCGCGCC 540  
Db 482 GGGGGCGGAGGTGCTGCTGACGTGCGCGCGCGGGGCGCTCCCGAGCGCCACACTGACT 541  
QY 541 GCGTCCGCTGGAGAGCTCGGCGAGGCAAGTGGCTGCGCATTCGGGCGCGCGCGCGCG 600  
Db 542 GGGAGAGGAGCGGATCGCCCTTGACGCAAGTGTGGGACAGCAGCCACTTCGCGCTCCAG 601  
QY 601 GCGACGCGGCACTTACAGGTCCGCGCGCGGAGAACCGCTGGGCG- - -GCTGCCAGCGCGCG 658

Db 602 CGGGCGCGCGGAGGACGGCCCGCGCGAGCTGGCACTCGCATCTCGCGCTCGGC 661  
QY 659 CGCGCGCTAGTG- - -GTGACTCGGACCGCGCGGACACGGCCAGCGCGCGGACCTCCA 717  
Db 662 TCCCGGATTCGGCGCTTACGTGTGCACGCCGCCAACCGCCACGGCCACCGCAGGCGG 721  
QY 718 CGCGCGCGCTCTCTGCGCACCTTGACAGCGCGCGCGAGGCTTATGCGCGCGAGGCGCGCC 777  
Db 722 GGGCGCTGCTCCAGGTGACACCGCCCGGAGAGCGCGCGCGGACCGCGAGGCGCC 781  
QY 778 CGCGCTC- - - - -ACGCGCCAGCACCGGACCGCGCACCTGACGGTGAATGAGGCA 828  
Db 782 CGCGCGCGGTGGTGGAGCGCGCTCAAGTGCGCCCTTAAGACCTTGGGTGAACGAGGCA 841  
QY 829 AGCAGCGCGGCTCAGCTGTACGTGACCGCGCGAGCGCCCAAGCGCGAGCGGTGTGGAAGA 888  
Db 842 AGCAGCGCAAGTTCGCTGCTACGTGTGTCGCAAGCGCGCGAGATCGAATGGCACT 901  
QY 889 AGGAGCGCCAGTGTGTACCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGAGGAGA 948  
Db 902 GGGAGGCGCGCGCGTGTCTCCGCGACCGCGCGCGCTCATGTACCGCGACCGCGCGCG 961  
QY 949 ACTTCGTGCTCAAGATCCTTCTTCTGCAAGCAGTCCGCGCGCGCGCTTACACCTTGCACGG 1008  
Db 962 GTTCTGCTCAGGTGCTTTTACTGCCAGGCGCAAGGATCGTGGGCTCTACGTCTGCGCG 1021  
QY 1009 COTCAACCTCGTGGCGCAGACCTACAGCTCTGCTGTGTGTAGTGGCGAGCGCGCGG 1068  
Db 1022 CGCGCAACTCGCGCGCGCGCAGACGCTCAGTGCCTGCGTGCAGCTGCAAGTGAAGTACGCGG 1081

RESULT 6  
AAS28907  
ID AAS28907 standard; DNA; 12415 BP.  
XX  
AC AAS28907;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immunoglobulin encoding genomic DNA SEQ ID No 269.  
KW Immunoglobulin; signal transduction pathway protein; cancer; ds;  
KW antisense therapy; gene therapy; neurological disorder; renal disorder;  
KW cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;  
KW reproductive disorder; immune system disorder; proliferative disorder;  
KW muscular disorder.  
XX  
OS Homo sapiens.  
XX  
XX WO200155315-A2.  
XX  
XX 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001326.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216860P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0217496P.  
PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225288P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225475P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 06-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 08-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234977P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0234984P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236328P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0244647P.  
PR 08-NOV-2000; 2000US-0244647P.  
PR 08-NOV-2000; 2000US-0244647P.

PR 08-NOV-2000; 2000US-0245476P.  
PR 08-NOV-2000; 2000US-0245477P.  
PR 08-NOV-2000; 2000US-0245478P.  
PR 08-NOV-2000; 2000US-0245523P.  
PR 08-NOV-2000; 2000US-0245524P.  
PR 08-NOV-2000; 2000US-0245525P.  
PR 08-NOV-2000; 2000US-0245526P.  
PR 08-NOV-2000; 2000US-0245527P.  
PR 08-NOV-2000; 2000US-0245528P.  
PR 08-NOV-2000; 2000US-0245532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249298P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 03-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WI; 2001-457725/49.  
DR  
XX Isolated novel immunoglobulin polypeptide for monitoring the presence and  
PT progression of diseases and for diagnosis.  
XX  
XX Claim 1; SEQ ID NO 269; 551bp; English.  
PS  
XX Sequences AAS28878-AAS28926 represent genomic DNA molecules which encode  
CC the immunoglobulin polypeptides of the invention. The polynucleotides and  
CC polypeptides can be used to diagnose a pathological condition or a  
CC susceptibility to a pathological condition in a subject by determining  
CC the presence or absence of a mutation in a DNA sequence or determining  
CC the presence or amount of expression of the protein. Alternatively the  
CC identification of a binding partner to a sequence allows determination of  
CC changes in protein activity. The sequences can be used as research tools  
CC for receptors or other signal transduction pathway proteins that interact  
CC with the polypeptides of the invention and can be used to treat, prevent  
CC or diagnose various types of disorders such as neurological disorders,  
CC cardiovascular disorders, gastrointestinal disorders, reproductive  
CC disorders, immune system disorders, renal disorders, muscular disorders,  
CC pulmonary disorders, proliferative disorders and cancer. Note: The  
CC sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 12415 BP; 2667 A; 3521 C; 3458 G; 2768 T; 0 U; 1 Other;

```
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231124P.
PR 08-SEP-2000; 2000US-023113P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-023237P.
PR 14-SEP-2000; 2000US-023237P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234223P.
PR 25-SEP-2000; 2000US-023497P.
PR 25-SEP-2000; 2000US-023498P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-024021P.
PR 02-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246603P.
PR 08-NOV-2000; 2000US-0246603P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249218P.
PR 06-SEP-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249247P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
XX used in preventing, treating or ameliorating a medical condition.
XX
XX Disclosure; SEQ ID NO 6044; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention
XX
XX Sequence 12415 BP; 2667 A; 3521 C; 3458 G; 2768 T; 0 U; 1 Other;
SQ
Query Match 18.5%; Score 197.6; DB 4; Length 12415;
Best Local Similarity 51.7%; Pred. No. 4,7e-23;
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;
QY 7 CCAGAGCCACACTCCGCGCTGCCAGAGCCGCCATAGAGAGGTCCTCCGCGGCA 66
Db 2 CCTCCGACCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 61
QY 67 CCGTCATGGATCAGCCACAGTTTCAGCGGGGCGCGCGCTTCTACCCCGCCAGGCT 126
Db 62 TGAAGCGGAGCTCGGGGATCAGGGAGCGCCCGCTTCTCCGCTTCCGCGGCTG 121
QY 127 TCGTGTGTGGTGGCAAGGACGCCACCTCAGCTGCCAGATCGTGGTAATCCACGC 186
Db 122 TCGGGGTGTAAGTGGCGCGGAGCGAGCTCAAGTGGTGTCTTGGGGAGCGCCGC 181
QY 187 CACAGGTGAGCTGGGAGAGGACACAGCAGCCGGTGACGGCGCGCGCTTCCGCTCG 246
Db 182 CTGTAGTGTGGTGGAGAGGGCGGCGAGCTGGCGCTCGGAACGCTTCCAGCTTC 241
QY 247 CCAGAGCGGAGCTCTTACCGCTTCACTATCTGTGACCTGGCGTGGCGAGCTGGGC 306
Db 242 CGCGGAGCGCGGAGCAGCGCTCTGTCTACCGCGCAGCTGCGCCACCGCGCGGG 301
QY 307 AATACGTGTCCGCGCGCAATGCCATAGGAGGCGCTTGGTGGGCTGCAGG 366
Db 302 TCTACGTGTCCGCGCGCGGAGCGGCGGAGGCTTACGGGGGCGCGCTTCCACCG 361
QY 367 TGGACCGGAGCGCGCTTCCGCGGAGCAGCGCGCGCTTCTGTGGGCGGCGGCTTCC 426
Db 362 TGCTGGAGCGCGCGCTTCCGCGGAGCTGAGCGCGCGCGCTTCCGCTTCCGCT 421
QY 427 TCCGCGTGGCGAGGCTCAGAGGCGGCTTCCGCTTCCGCTTCCGCTTCCGCT 480
```

Db 422 CGGGTCCGGAGGGGCCCCCGCTTCTTCACGGGGCTTCGATCCAGTGGGTGTCG 481  
QY 481 CGAGCCCGCAGTGAAGTGGTCCAGGACGGGGCGCGCTGGGTGAGCCGACGGCCCCC 540  
Db 482 GGGGGCGAGTGGTGTGCTGACGTCCGGGGCGGGGGCTCCCGAGGCCACACTGTACT 541  
QY 541 CGGTCCGCTGAGAGAGTCCGGAGGCAATGTCGCTGCGCATTCGGGGCGGGCGCCG 600  
Db 542 GGGAGAAGCAGGGATGCCCTTGAACGAAGTGTGGACAGCAGCCACTTCGCGTCCAGC 601  
QY 601 CGCAGCGGGCACTTACAGGTCCGCGCGAGAACCCGCTGGGC -GCTGCCAGCGCGC 658  
Db 602 CGGGCGCGCGAGGACGGCCCCGGCGGAGCTGGCACTGCGCATCTGGCGCTCGC 661  
QY 659 CGGGCGCTAGTG-GTGGACTCGACGCGCGGACACGGGCGAGCCCGCGGACTTCA 717  
Db 662 TCGCGGATTCGGGCTTACGTGTGCCACGCCCGCAACGGCGACGGCGAGCGG 721  
QY 718 CGGGCGGCTCTGCGCACTGCGAGCGCGGCGGCGGAGGTATGCGCGCGAGGGCGCC 777  
Db 722 GGGGCTGCTCCAGTGCACAGCCCCCGGAGAGCCCGCGGAGCCCGGAGAGGCC 781  
QY 778 CCGGCTC-----ACGGCCAGCACCGGCGAGCTGCACGGTGAAGGCA 828  
Db 782 CCGCGCGGTGGAGCGCTCAAGTGGCGGCTAAGACCTTCTGGGTGAACGAGGCA 841  
QY 829 AGCAGCGCGCTCAGCTGTACGTGACCGGCGAGCCAGCGGAGCGGTGGAAGA 888  
Db 842 AGCAGCGCAAGTTCGGTGTACGTGATGGCAAGCCGAGCGGAGATCGAATGCACT 901  
QY 889 AGGACGGCCAGCTGCTGACGAGGGCGCGGCGGCGGCTGTACGAGGAGCGCAGAGA 948  
Db 902 GGGAGGCGCGCGCTGCTCCCGACCGCGCGCTCATGTACCGGACCGGAGCGG 961  
QY 949 ACTTGTGCTCAAGATCTCTTTCGAGAGTCCGAGCGCGGCTTACCTGCAAGG 1008  
Db 962 GCTTCGTGCTCAAGTGTCTTACTGCCAGGCGGAGGATCGTGGGCTCTAGCTCGGCGG 1021  
QY 1009 GGTCAACCTGTCGGGCGAGACTTACAGCTGTGTGTCGTCGTCGAGCGGCGGCGG 1068  
Db 1022 CGCGCAACTCGGGCGGCGAGCGCTCAGTCCGCTGACGTGACGTGAAGGTACGCGG 1081

## RESULT 8

ADB31748

ID ADB31748 standard; DNA; 12415 BP.

AC ADB31748;

XX DT 04-DEC-2003 (first entry)

XX DE Human novel protein DNA SEQ ID NO 269.

XX KW Gene therapy; human; immunoglobulin; cancer; lung cancer; leukaemia;

XX KW ovarian cancer; epithelial cancer; hyperplasia; Gaucher's disease; AIDS;

XX KW arrhythmia; cardiac oedema; ischaemia; pneumonia; cystic fibrosis;

XX KW asthma; sarcoidosis; rhinitis; anaemia; inflammation; sinusitis;

XX KW chronic obstructive pulmonary disease; infectious disease; gene; ds.

XX OS Homo sapiens.

XX PN US2003077606-A1.

XX PD 24-APR-2003.

XX PF 07-MAR-2002; 2002US-00091438.

XX PR 31-JAN-2000; 2000US-0179065P.

XX PR 04-FEB-2000; 2000US-0180628P.

XX PR 24-FEB-2000; 2000US-0184664P.

XX PR 02-MAR-2000; 2000US-0186350P.

XX PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226689P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0232403P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 21-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234233P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234977P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.

13-OCT-2000; 2000US-0239937P.  
 20-OCT-2000; 2000US-0240960P.  
 20-OCT-2000; 2000US-0241221P.  
 20-OCT-2000; 2000US-0241785P.  
 20-OCT-2000; 2000US-0241786P.  
 20-OCT-2000; 2000US-0241787P.  
 20-OCT-2000; 2000US-0241808P.  
 20-OCT-2000; 2000US-0241809P.  
 20-OCT-2000; 2000US-0241826P.  
 01-NOV-2000; 2000US-0244617P.  
 08-NOV-2000; 2000US-0246474P.  
 08-NOV-2000; 2000US-0246475P.  
 08-NOV-2000; 2000US-0246476P.  
 08-NOV-2000; 2000US-0246477P.  
 08-NOV-2000; 2000US-0246478P.  
 08-NOV-2000; 2000US-0246523P.  
 08-NOV-2000; 2000US-0246524P.  
 08-NOV-2000; 2000US-0246525P.  
 08-NOV-2000; 2000US-0246526P.  
 08-NOV-2000; 2000US-0246527P.  
 08-NOV-2000; 2000US-0246528P.  
 08-NOV-2000; 2000US-0246532P.  
 08-NOV-2000; 2000US-0246609P.  
 08-NOV-2000; 2000US-0246610P.  
 08-NOV-2000; 2000US-0246611P.  
 08-NOV-2000; 2000US-0246613P.  
 17-NOV-2000; 2000US-0249207P.  
 17-NOV-2000; 2000US-0249208P.  
 17-NOV-2000; 2000US-0249209P.  
 17-NOV-2000; 2000US-0249210P.  
 17-NOV-2000; 2000US-0249211P.  
 17-NOV-2000; 2000US-0249212P.  
 17-NOV-2000; 2000US-0249213P.  
 17-NOV-2000; 2000US-0249214P.  
 17-NOV-2000; 2000US-0249215P.  
 17-NOV-2000; 2000US-0249216P.  
 17-NOV-2000; 2000US-0249217P.  
 17-NOV-2000; 2000US-0249218P.  
 17-NOV-2000; 2000US-0249244P.  
 17-NOV-2000; 2000US-0249245P.  
 17-NOV-2000; 2000US-0249264P.  
 17-NOV-2000; 2000US-0249265P.  
 17-NOV-2000; 2000US-0249297P.  
 17-NOV-2000; 2000US-0249299P.  
 17-NOV-2000; 2000US-0249300P.  
 01-DEC-2000; 2000US-0250160P.  
 05-DEC-2000; 2000US-0251030P.  
 08-DEC-2000; 2000US-0251988P.  
 08-DEC-2000; 2000US-0256719P.  
 08-DEC-2000; 2000US-0251479P.  
 08-DEC-2000; 2000US-0251856P.  
 08-DEC-2000; 2000US-0251868P.  
 08-DEC-2000; 2000US-0251869P.  
 08-DEC-2000; 2000US-0251989P.  
 08-DEC-2000; 2000US-0251990P.  
 11-DEC-2000; 2000US-0254097P.  
 05-JAN-2001; 2001US-0259678P.  
 17-JAN-2001; 2001US-00764879.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Ruben SM, Barash SC;  
 WPI; 2003-615993/58.  
 New human immunoglobulin superfamily of polypeptide and genes, useful for treating, preventing or diagnosing e.g. cancers (lung cancer, leukemia, ovarian epithelial cancer, etc.), hyperplasia, ischemia, pneumonia or AIDS.  
 Disclosure; SEQ ID NO 269; 213pp; English.

RESIST. 9

AAK89083  
ID AAK89083 standard: DNA: 20565 BP.

05-NOV-2001 (first entry)

XX  
XX

xx Homo sapiens.

PD 02-AUG-2001.

PR 31-JAN-2000; 2000US-0179065P.

PR 02-MAR-2000; 2000US-0186350P.

PR 17-MAR-2000; 2000US-0190076P.

PR 19-MAY-2000; 2000US-0205515P.

PR 28-JUN-2000; 2000US-0214886P.

PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.

PR 26-JUL-2000: 2000US-0220963P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P

14-AUG-2000; 2000US-0225759P  
14-AUG-2000; 2000US-0225759P  
14-AUG-2000; 2000US-0225759P

[illegible]

PR	18-AUG-2000;	2000US-02263798P
PR	22-AUG-2000;	2000US-02266816P
PR	22-AUG-2000;	2000US-02266868P
PR	23-AUG-2000;	2000US-02271822P
PR	23-AUG-2000;	2000US-02277009P
PR	30-AUG-2000;	2000US-02289242P
PR	31-AUG-2000;	2000US-02289287P
PR	01-SEP-2000;	2000US-02293433P
PR	01-SEP-2000;	2000US-02293444P
PR	01-SEP-2000;	2000US-02293454P
PR	05-SEP-2000;	2000US-02295039P
PR	05-SEP-2000;	2000US-02295109P
PR	06-SEP-2000;	2000US-02304372P
PR	06-SEP-2000;	2000US-02304381P
PR	08-SEP-2000;	2000US-02304342P
PR	08-SEP-2000;	2000US-02312433P
PR	08-SEP-2000;	2000US-02312443P
PR	08-SEP-2000;	2000US-02314133P
PR	08-SEP-2000;	2000US-02314143P
PR	08-SEP-2000;	2000US-02320801P
PR	12-SEP-2000;	2000US-02331968P
PR	14-SEP-2000;	2000US-02332977P
PR	14-SEP-2000;	2000US-02332988P
PR	14-SEP-2000;	2000US-02333999P
PR	14-SEP-2000;	2000US-02334001P
PR	14-SEP-2000;	2000US-02336033P
PR	14-SEP-2000;	2000US-02336044P
PR	14-SEP-2000;	2000US-02336055P
PR	21-SEP-2000;	2000US-02342767P
PR	21-SEP-2000;	2000US-02342777P
PR	25-SEP-2000;	2000US-02345972P
PR	25-SEP-2000;	2000US-02345988P
PR	26-SEP-2000;	2000US-02334844P
PR	27-SEP-2000;	2000US-02335634P
PR	27-SEP-2000;	2000US-02335637P
PR	29-SEP-2000;	2000US-02336327P
PR	29-SEP-2000;	2000US-02336376P
PR	29-SEP-2000;	2000US-02336378P
PR	29-SEP-2000;	2000US-02336369P
PR	29-SEP-2000;	2000US-02336702P
PR	02-OCT-2000;	2000US-02336802P
PR	02-OCT-2000;	2000US-02370377P
PR	02-OCT-2000;	2000US-02370388P
PR	02-OCT-2000;	2000US-02370393P
PR	02-OCT-2000;	2000US-02370409P
PR	13-OCT-2000;	2000US-02399355P
PR	13-OCT-2000;	2000US-02399377P
PR	20-OCT-2000;	2000US-02409606P
PR	20-OCT-2000;	2000US-02412121P
PR	20-OCT-2000;	2000US-02417855P
PR	20-OCT-2000;	2000US-02417866P
PR	20-OCT-2000;	2000US-02417876P
PR	20-OCT-2000;	2000US-02418068P
PR	20-OCT-2000;	2000US-02418099P
PR	20-OCT-2000;	2000US-02418262P
PR	01-NOV-2000;	2000US-02446174P
PR	08-NOV-2000;	2000US-02464747P
PR	08-NOV-2000;	2000US-02464757P
PR	08-NOV-2000;	2000US-02464767P
PR	08-NOV-2000;	2000US-02464776P
PR	08-NOV-2000;	2000US-02464788P
PR	08-NOV-2000;	2000US-02465232P
PR	08-NOV-2000;	2000US-02465242P
PR	08-NOV-2000;	2000US-02465252P
PR	08-NOV-2000;	2000US-02465262P
PR	08-NOV-2000;	2000US-02465272P
PR	08-NOV-2000;	2000US-02465284P
PR	08-NOV-2000;	2000US-02465322P
PR	08-NOV-2000;	2000US-02466092P
PR	08-NOV-2000;	2000US-02466101P
PR	08-NOV-2000;	2000US-02466116P





KW reproductive disorder; immune system disorder; proliferative disorder;  
muscular disorder.  
XX  
XX Homo sapiens.  
XX WO200155315-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US001326.  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 14-AUG-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228944P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231143P.  
PR 08-SEP-2000; 2000US-0231144P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235838P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251866P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.

FR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-457725/49.  
DR Isolated novel immunoglobulin polypeptide for monitoring the presence and  
XX progression of diseases and for diagnosis.  
PT Claim 1; SEQ ID NO 270; 551pp; English.  
XX  
XX Sequences AAS28978-AAS28926 represent genomic DNA molecules which encode  
CC the immunoglobulin polypeptides of the invention. The polynucleotides and  
CC polypeptides can be used to diagnose a pathological condition or a  
CC susceptibility to a pathological condition in a subject by determining  
CC the presence or absence of a mutation in a DNA sequence or determining  
CC the presence or amount of expression of the protein. Alternatively the  
CC identification of a binding partner to a sequence allows determination of  
CC changes in protein activity. The sequences can be used as research tools  
CC for receptors or other signal transduction pathway proteins that interact  
CC with the polypeptides of the invention and can be used to treat, prevent  
CC or diagnose various types of disorders such as neurological disorders,  
CC cardiovascular disorders, gastrointestinal disorders, reproductive  
CC disorders, immune system disorders, renal disorders, muscular disorders,  
CC pulmonary disorders, proliferative disorders and cancer. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 20565 BP; 4414 A; 5760 C; 6155 G; 4236 T; 0 U; 0 Other;

Query Match 18.5%; Score 197.6; DB 4; Length 20565;  
Best Local Similarity 51.7%; Pred. No. 4.4e-23;  
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

QY 7 CCAGCAGCCACACTCGCGCGCTGCCAGAGCCGCCATAGAGAGAGTCCCGCCGCCA 66  
DB 20564 CCTTCGCCACCCCCCTGCCCCCACCCTGCTCGCGCTGCGAGCGCTGCGCGCCGA 20505

QY 67 CCGTCATGATPACGCCAGTTCAGCGGGCGCCCGCTTCTTCCACCGGCCCAAGCCCT 126  
DB 20504 TGAAGCGAGCTCGGGGATCAGGGAGCCCGCTGCTTCTCGCTTCCCGGCGCTG 20445

QY 127 TCGTGGTGTGCTGGGAGAGCCAGCCACCTCAGCTCCAGATCGTGGTAAATCCACGC 186  
DB 20444 TCGCGGTGGTAAGTGGCGCCGAGCGCCAGCTCAAGTGGTGGTCTTGGGGAGCCCGCC 20385

QY 187 CACAGGTGAGCTGGGAGAGACACAGCAGCGCTGACCGCGCGCGCTTCCGTCTGG 246  
DB 20384 CTGTAGTGGTGGGAGAGCGCGGCGAGCTGCGGCTCGAAGCGCTGAGCTTCC 20325

QY 247 CCGAGGAGCGGACCTTACCGCTCTACTATCTGTGAGCTGGCGCTGGGCGAGTGGGC 306  
DB 20324 CGGCGGAGCGGCGGAGCAGCGCTGTGTGTGACCGCGCAGCTGCCACCGAGCGGGG 20265

QY 307 AATAGCTGTGCGCGCGGCAATGCCATAGCGAGGCTTTCGCTGCGTGGCGCTGACG 366  
DB 20264 TCTAGTGTGCGCGCCCGCAACCGCGCGCGGAGGCTTACGCGCGCGCGCTACCG 20205

QY 367 TGGACGCGGAGCGCGCTGCGCGCGAGCGCGCGCACTTCTCTCTCGGCGCCACGTCCA 426  
DB 20204 TGCTGGAGCGCGCGCTTCCAGCCCGAGCTGACGCGCGCGCGCGCTGCGCATCGC 20145

QY 427 TCCCGCTGCGGAGGCTCAGAGCGCCACCTTCCGCTCCG-----CGTGGTGGCTCCC 480  
DB 20144 CGGGTCCGGGAGGCGCGCGCGCTTCTTCTTCAAGGGGCTTCTGATCCAGTGGTGTGTC 20085

QY 481 CGAGCGCGGAGTGAAGTGGTCCAGAGCAGCGCGCGCGCTGGGTGAGCGCGAGCGCGCCCC 540  
DB 20084 GGGGGCGGAGTGGTGTGCTGACGTGCCGGGCGGGGGCTTCCCGAGCGCCACACTGTACT 20025

QY 541 GCGTCGGCTGGAGAGTCTGCGCGAGGCAAGTGCCTCGGATTCGGCGCGCGCGCGCGC 600  
DB 20024 CGGAGAAGGACGGGATGCGCTTGGACGAAGTGTGGGACAGAGCCACTTTCGCTTCCAGC 19965

QY 601 GCGAGCGCGGACTTACGAGGTCCCGCGCGGAGAACCCGCTGGC--GCTGCCAGCGCGC 658  
DB 19964 CGGGCGCGCGGAGGCGCGCGCGCGCGAGCTGGCATCTTGGCGGTTCGCG 19905

QY 659 CGCGCGCTGTAGTG--GTGACTCTGCGCGCGGAGCACGCGCGCGCGCGCGCGCGCTCCA 717  
DB 19904 TCGCGGATTCGCGCGTCTACGTGTGCCACGCCCGCAACGCGCACGCGCGCGCGG 19845

QY 718 CGGCGCGCTCTCTGCGCACCTGCGAGCGCGCGCGCGAGCTATGCGCGCGCGCGCGCGC 777  
DB 19844 GGGCGCTGTCTCAGGTGCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 19785

QY 778 CGCGCTC-----ACCGC 828  
DB 19784 CGCGCGCGGTGTGGGCGCGCTCAAGTGGCGCGCTTCTGGGTGAACGAGGCA 19725

QY 829 AGCAGCGCGCGCTCAGCTGTCTACGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 888  
DB 19724 AGCAGC 19665

QY 889 AGGAGCGCGCGCTGTGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 948  
DB 19664 GGGAGCGCGCGCGCTGTCTTCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 19605

QY 949 ACTTGTGTCTCAAGATCTTCTTCTTCAAGCAGTTCGGAACCGCGCGCTTACACTTGCACG 1008  
DB 19604 GCTTGTGTCTCAAGGTCTTACTGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGC 19545

QY 1009 GCTCCACCTCTGGC 1068  
DB 19544 CGCGCAACTCGC 19485

RESULT 11  
AAL03357  
ID AAL03357 standard; DNA; 20565 BP.  
XX  
AC AAL03357;  
XX  
DT 21-NOV-2001 (first entry)  
XX  
DE Human reproductive system related antigen DNA SEQ ID NO: 6045.  
XX  
KW Human; reproductive system related antigen; reproductive system disorder;  
XX cancer; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200155320-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001339.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189974P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0190123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0231428P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0232081P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234233P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0234984P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236377P.  
PR 29-SEP-2000; 2000US-0236378P.  
PR 29-SEP-2000; 2000US-0236386P.  
PR 29-SEP-2000; 2000US-0236389P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241755P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.

PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.

Disclosure; SEQ ID NO 6045; 1297pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention

Sequence 20565 BP; 4236 A; 6155 C; 5760 G; 4414 T; 0 U; 0 Other;

Query Match

Best Local Similarity 18.5%; Score 197.6; DB 4; Length 20565;

Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

QY

7 CCAGCAGCCCACTCCGGCGGTGCCAGAGCCCATAGAGAGGTCCCGCCGCCA 66

Db 2 CCCTCCGACCCCCCTGCCCCCACCCTTCCGCGCTGCAGCGGTCCGCGCGCGCA 61  
QY 67 CCGTCATGATACCCACAGTTACAGCGGGCGCCCGCTTCTCACCCGCGCCAAAGCCT 126  
Db 62 TGAAGCGAGCTCGGGGATACAGGGAGCCCCCGTGTCTCTCGGCTTCCCGCGCCTG 121  
QY 127 TCGTGTGTGTGGTGGGAAGAACCCACCTTCACTGCTCCAGATCGTGGGTAATCCACGC 186  
Db 122 TCGGGTGGTAAGTGGCGCGAGCCGAGCTCAAGTGCCTGGTCTTGGGGAGCCCGCGC 181  
QY 187 CACAGTGTAGTGGGAGAACACAGACCGGTGAGCGCGCGCGCTTCCGCTCG 246  
Db 182 CTGTAGTGTGTGGGAAGAGCGCGGAGAGCTGGCGGCTTGGAAAGCCTGAGCTTC 241  
QY 247 CCCAGGACGGGACCTTACCGCTCTACTATCTCTGACCTTGGCGCTGGGCGACAGTGGC 306  
Db 242 CGCGGACGGCGGAGACAGCGCTGCTGTGACCGCGCACTGCCACACGACGCGGGG 301  
QY 307 AATAGTGTGCGCGCGCGGCAATCCATAGCGAGGCTTCTGCTGCGCTGGGCTGCAGG 366  
Db 302 TCTACGTGTGCGCGCGCGCAACCGCGCGGAGGCTTACGGCGCGCGCGCTACCG 361  
QY 367 TGGACGGGAGGCGCGTGCAGCGAGCGCGCGCACTTCTGCTCGCGCCACAGTCCA 426  
Db 362 TGCTGGAGCGCGCGCTTCCGACCCCGAGCTGACCGCGCGCGCGCTGCCATCGC 421  
QY 427 TCCCGGTGCGAGGGTCAAGGCGCACTTCCGCTGCGG-----CGTGGGTGGCTCC 480  
Db 422 CGGGTTCGGGAGGCGCGCGCTTCTCTCACGGGCGCTCGATCCAGTGGGTGCTG 481  
QY 481 CGAGCGCGGAGTGAAGTGTCCAGAGCGCGCGCGCTGGGTGAGCGCGCGCGCC 540  
Db 482 GGGGGCGGAGGTGTGCTGAGCTGTCGCGCGCGCGCGCGCTCCCGAGCCACACTGTACT 541  
QY 541 CGTGTGCGGTGAGAGCTCGCGAGGCAAGTGGCGCTGCGCATTCGGGCGCGCGCGCG 600  
Db 542 GGGAGAAGACGGGATGCGCTGACAGCAAGTGTGGACAGCAGCACTTCGCGCTCCAGC 601  
QY 601 CGAGCGCGGCACTTACAGAGTCCGCGCGAGACCCGCTGGGCG--CGTCCGACGCGCG 658  
Db 602 CGGGCGCGCGAGACCGCGCGCGCGCGAGCTTGGCACTTCTGCGCGCTCGG 661  
QY 659 CGCGCGCTAGTG--GTGACCTCGAGCGCGCGGACACCGCGCGCGCGCGCGCGCGCTCCA 717  
Db 662 TGCGGATTCGGCGCTTACGTGTGCGACGCGCGCGCGCGCGCGCGCGCGCGG 721  
QY 718 CGCGCGCGCTTCTGCGCACTGACAGCGCGCGCGGAGGCTATGGCGCGCGAGGCGCGC 777  
Db 722 GGGCGCTGCTCAGGTGACCAAGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCG 781  
QY 778 CGCGCTC-----ACCGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 828  
Db 782 CGCGCGGTGTGAGCGCGCTCAAGTCCGCGCTTACAGCTTGTGGTGAACGAGGCA 841  
QY 829 AGCAGCGCGCTCAGCTGCTACGTGACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 888  
Db 842 AGCAGCGCAAGTTCGCTGCTACGTGATGGGCAAGCGCGCGCGCGCGCGCGCGCGCG 901  
QY 889 AGGAGCGCGAGTGTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 948  
Db 902 GGGAGGGCGCGCGCTGTCCGCGACCGCGCGCGCTCATGTACCGCGCGCGCGCGCGCG 961  
QY 949 ACTTCGCTCAGATCTCTTCTGCAAGCAAGTCCGAGCGCGCGCGCGCTTACACCTGCA 1008  
Db 962 GCTTCGTGCTCAAGTGTCTTACTGCGAGCGCAAGGATCGTGGGCTCTACGCTCTGCG 1021  
QY 1009 CGTCCAACTCGTGGGCGAGACTACAGCTCTGTGCTGTGTGCTGAGCGCGCGCGCGG 1068  
Db 1022 CGCGCAACTCGCGCGCGCGCGCGCTCAGTGGCGGTGCGAGCTGCACTGAAAGTACGCG 1081

RESULT 12

ADB31749/c  
ID ADB31749 standard; DNA; 20565 BP.  
XX  
AC ADB31749;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Human novel protein DNA SEQ ID NO 270.  
XX  
KW gene therapy; human; immunoglobulin; cancer; lung cancer; leukaemia;  
ovarian cancer; epithelial cancer; hyperplasia; Gaucher's disease; AIDS;  
arrhythmia; cardiac oedema; ischaemia; pneumonia; cystic fibrosis;  
asthma; sarcoidosis; rhinitis; anaemia; inflammation; sinusitis;  
chronic obstructive pulmonary disease; infectious disease; gene; ds.

OS Homo sapiens.  
XX  
PN US2003077606-A1.  
XX  
PD 24-APR-2003.  
XX  
PF 07-MAR-2002; 2002US-00091438.  
XX  
31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216547P.  
PR 07-JUL-2000; 2000US-0218880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231444P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.



RESULT 13	
ADD14722	
ID	ADD14722 standard; cDNA; 5382 BP.
XX	
XX	
AC	ADD14722;
XX	
DT	01-JAN-2004 (first entry)
XX	
DE	Human src biomarker polynucleotide SEQ ID NO:116.
XX	
KW	predictor set; protein tyrosine kinase activity modulator;
KW	protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
KW	gene therapy; drug sensitivity; genetic profile; cancer; human; gene; ss.
XX	
OS	Homo sapiens.
XX	
FN	WO2003062395-A2.
XX	
XX	
PD	31-JUL-2003.
XX	
XX	
PF	17-JAN-2003; 2003WO-US001981.
XX	
PR	18-JAN-2002; 2002US-0350061P.
XX	
XX	

125	GTCTACGTGTGCCACGCCCGCAACCGACACGGCCACGGCGACGCGCGGGGGCGTGTCTCCAG	154	Db
732	GGCGACCTTCAGCGCGGGCGGAGGACTATTCGCGCGCGAGGGCGCCCCCGCCCTC	784	QY
185	GTGCACCAACCCCGGAGAGCCCGCGCGGACCCCGACGAGGCCCGCCGCGCGTGGTG	244	Db
785	--ACCGCCGAGCACCGGCACGGCACCTTCACAGTGTACTGAAGGCAAGCACGCGCGCCTC	842	QY
245	GAGCGCTCAAGTGCGGCGCTTAAGACCTTCTGGGTGAACGAGGGCAAGCACGCCAAGTTC	304	Db
843	AGCTGCTAGTGACCGCGAGCCCAAGCCCGAGACGGTGTGGAAGAGACGCGCCAGCTG	902	QY
305	CGCTGCTACGTATGGCAAGCCCGAGCCCGAGATCGAATGGCACTGGGAGGGCCCGCCG	364	Db
903	GTGACCGAGGGCGCGCCACGTGTGTACAGGAGCGCGCAGGAGAACTCTGTGTCTCAAG	962	QY
365	CTGTCTCCGGACCGCGCGCGCTCATGTACCGCGACCGCGGCTCTGTGTCTCAAG	424	Db
963	ATCTCTTTCTGAAGAGTGCAGCCGCGGCTCTTACACTGTGCACGGCGTCCCACTCGTG	1022	QY
425	GTGCTCTACTGCCAGGCCAAGGATCGTGGGCTCTACGTCTGCGCCGCGGCACTCGGCG	484	Db
1023	GGCCAGACCTACAGCTCTGTGTGGTCTGTAGTGTGCGGAGGCC	1064	QY





Db 6915 CCCCCTGGACGGACACCTGACGCGCAGACCCCGCGTTCGCCCTGCAGACCGCGCT 6974  
QY 404 CTTCTCTGTGCGGCCACAGTCCATCCGGTGTGCGAGGGCTCAGAGGCCACTTCCGGTG 463  
Db 6975 GSCCCCCAGCTGGCGGCTTCGGCTCGCGCCGACCGGGTGGCGGCCACTCGGTGG 7034  
QY 464 CCGGTGTGGTCTCCCGAGGCCGGCAGTGTGTCGAAGGACCGCGCGCGCTGGG 523  
Db 7035 CGAGATCGCGCGCGCGACGTGCGCGGCGCTTCGCCCTCGACGCGCGTGGTGGT 7094  
QY 524 TGAGCCGACGCGCCCGCGTGTGGTGTGAGAGCTCGCGAGGCAAGTGGCTGCGCAT 583  
Db 7095 CCGCAGCGCGCGTGTCTATCGCGACCTGCGACCGGGCGCATGTCGCCGTGAGGT 7154  
QY 584 TCGGGCGCGCGCGCGCGACCGCGCACTTA-----CGAGGTCCGCGCGAGAA 634  
Db 7155 CACCGAGAGGAGCGCTGGACGCGATCGCGACCTGCCCGTCCGCGTCCGCTGTCAA 7214  
QY 635 CCGGTGTGGCTGTCCAGCGCGCGCGCGCTAGTGTGACTCGGACGCGCGCGGACAC 694  
Db 7215 CCGCGCGAGGCGGTGCTCTCCGCGACCGCGAGCGGTGTCGGCGCTCGCGGCCG 7274  
QY 695 GGCACGCGCGCGCGCGACCTCCACGCGCGCGCTCCTGGCGCACCTGCAGCGCGCGCGCGA 754  
Db 7275 CCTCGCGCGCGCGCAAGCGCTCCCGTGGACCGGGGTTCCACTCGCCGACGTGA 7334  
QY 755 GGCTATGC---GCGCGAGGGCGCCCGCTTCACCGCCGACGCGGACCGGACCTG 811  
Db 7335 CCGGTGTGACGCGCTGCGCGCTGCGCCACCGCGACCCCGACACACCCCTGATCTC 7394  
QY 812 CACGTGTACTGAAGCAAGCACGCGCTCAGCTGTACGTGACCGCGCGAGCCCAAGCC 871  
Db 7395 CACCAACCACCGCGCGACGCTGACCCACGTCGCGCACACCACTGGGCCGACCGCGG 7454  
QY 872 CGAGACGTGTGGAAAGGACGCGCCAGTGTGACCGAGGCGCGCGCCACGTTGTGTA 931  
Db 7455 CGCGCGGTTCGCTTCGCGGACGCGCTGCGCGCCCTGGCGGCGACGCGCGTCAAGCGG 7514  
QY 932 CGAGACGCGCAGGAGAACTTGTGTCAAGATCCTTTCGACAGAGTGGACCGCGG 991  
Db 7515 CGTGAGCTGGGCGCGCGCGTGTACCCCGCACCGCGCGAGTCTTGGCGCGCGC 7574  
QY 992 CCTTACACCTGCACGCGGTCCAACTCGTGGGCCAGACCTACAGCTCTGTGTGTGT 1051  
Db 7575 GATCCCGCGCTGCGCGCGGACCAAGAGGTGTCGCTCGCTCACCGCGTGGCGT 7634  
QY 1052 AGTGGCGGAGCCG 1065  
Db 7635 GGTGCACGCGCG 7648

Search completed: September 19, 2004, 15:42:33  
Job time : 448.18 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 15:09:24 ; Search time 96.0922 Seconds  
(without alignments)  
6179.453 Million cell updates/sec

Title: US-10-077-130-4 COPY 1 1070

```

perfect score: 1070
Sequence: 1 tgccctaccagagacccacac.....tagtcgcgagacccgcggtt 1070

```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs. 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum MACCH 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*

```
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	68.6	6.4	1614	4	US-09-616-289-45
C 2	68.6	6.4	12425	4	Sequence 45, Appl
C 3	55	6.1	3375	4	Sequence 50, Appl
C 4	64.8	6.1	44377	2	Sequence 1395, Ap
C 5	64.8	6.1	44377	2	Sequence 7, Appl
C 6	62.6	5.9	8438	1	Sequence 1, Appl
C 7	62.6	5.9	4403765	3	Sequence 1, Appl
C 8	61.8	5.8	1191	3	Sequence 2, Appl
C 9	61.8	5.8	1389	1	Sequence 3, Appl
C 10	61.8	5.8	1797	1	Sequence 41, Appl
C 11	61.8	5.8	1839	4	Sequence 2, Appl
C 12	61.8	5.8	1839	4	Sequence 1, Appl
C 13	61.8	5.8	1869	2	Sequence 1, Appl
C 14	61.8	5.8	2760	2	Sequence 15, Appl
C 15	61.8	5.8	2760	2	Sequence 20, Appl
C 16	61.8	5.8	2760	3	Sequence 3, Appl
C 17	61.8	5.8	2760	3	Sequence 1, Appl
C 18	61.8	5.8	5727	5	Sequence 13, Appl
C 19	61.6	5.8	4403	2	Sequence 1, Appl
C 20	61.6	5.8	4403	2	Sequence 1, Appl
C 21	61.6	5.8	4403	3	Sequence 1, Appl
C 22	61.6	5.8	4403	5	Sequence 1, Appl
C 23	61.2	5.7	4257	2	Sequence 1, Appl
C 24	61.2	5.7	4257	3	Sequence 1, Appl
C 25	61.2	5.7	4257	3	Sequence 1, Appl
C 26	61	5.7	2561	4	Sequence 48, Appl
C 27	60.8	5.7	2185	1	Sequence 3, Appl

28	60.8	5.7	2185	2	US-08-365-310-3	Sequence 3, Appli
29	60.8	5.7	2185	3	US-08-351-742-3	Sequence 3, Appli
30	60.4	5.6	1029	4	US-09-352-591A-135	Sequence 135, App
31	60.4	5.6	2457	4	US-09-352-591A-120	Sequence 120, App
32	60.4	5.6	2707	2	US-08-709-874A-9	Sequence 9, Appli
33	60.4	5.6	2707	3	US-09-104-382-9	Sequence 9, Appli
34	60.4	5.6	2707	4	US-09-833-555-9	Sequence 9, Appli
35	60.2	5.6	2001	4	US-09-352-591A-3177	Sequence 3177, Ap
36	60.2	5.6	2012	1	US-08-235-838-13	Sequence 13, Appl
37	60.2	5.6	2012	1	US-08-235-838-15	Sequence 15, Appl
38	60.2	5.6	2012	2	US-08-465-473B-13	Sequence 13, Appl
39	60.2	5.6	2012	2	US-08-465-473B-15	Sequence 15, Appl
40	60.2	5.6	2304	4	US-09-352-591A-2790	Sequence 2790, Ap
41	60.2	5.6	2561	4	US-09-616-289-48	Sequence 48, Appl
42	59.8	5.6	77536	1	US-09-410-551B-1	Sequence 1, Appli
43	59	5.5	1187	1	US-08-440-854A-2	Sequence 2, Appli
44	58.6	5.5	1908	3	US-09-046-999-1	Sequence 1, Appli
45	58.6	5.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli

## ALIGNMENTS

## RESULT. T. 1

```

US-09-616-289-45/C
/ Sequence 45, Application US/09616289
/ Patent No. 6632923
/ GENERAL INFORMATION:
/ APPLICANT: Lees, Robert S.
/ APPLICANT: Lees, Ann M.
/ APPLICANT: Law, Simon W.
/ APPLICANT: Ariana, Anibal A.
/ TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
/ TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
/ TITLE OF INVENTION: ATHEROSCLEROSIS
/ FILE REFERENCE: 10797-004001
/ CURRENT APPLICATION NUMBER: US/09/616,289
/ CURRENT FILING DATE: 2000-07-14
/ PRIOR APPLICATION NUMBER: US 09/517,849
/ PRIOR FILING DATE: 2000-03-02
/ PRIOR APPLICATION NUMBER: US 08/979,608
/ PRIOR FILING DATE: 1997-11-26
/ PRIOR APPLICATION NUMBER: US 60/031,930
/ PRIOR FILING DATE: 1996-11-27
/ PRIOR APPLICATION NUMBER: US 60/048,547
/ PRIOR FILING DATE: 1997-06-03
/ NUMBER OF SEQ ID NOS: 53
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 45
/ LENGTH: 1614
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)...(1614)
US-09-616-289-45

```

Query Match	6.4%	Score 68.6;	DB 4;	Length 1614;
Best Local Similarity	45.7%;	Pred. No. 0.00029;		
Matches 274;	Conservative	0;	Mismatches 324;	Indels 1;
				Gaps 1;

238	TCGCTCTGGCCCAAGAGCGGGAACTCTACCGCTCACTATCTCTGACCTCGCGCTGGCGG	297
QY		
863	TCCAGACGGCTCTCAGACCGCGCGCTCTCTCTCAGCAGCCCCCTGCACGCGCGCGCGGGTT	804
DB		
298	ACAGTGGGCAATACGTGTGCGCGCGCGCAATGCCATAGCGAGGCGCTTCGTGCGGTGG	357
QY		
803	AGGCACCGCGCGCGCGCGCTGCCCGGAGTAGCGCACACTTCCCGCAGGCTCAG	744
DB		
358	GCCTGCAGGTGGACGCGAGCGCGCGTGCGCCGAGCAGCGCGCGCACTTCTCTGCTGCGCG	417
QY		
743	GGCGGCG	684
DB		





QY 996 TACACCTGCACGGCTCCAACTCGTGGGCGACACCTA 1033  
DB 24340 GACGACATCCCGGACGTCAGCGGCTGCGCGCGGCCGA 24377

RESULT 5  
US-08-804-198-1  
; Sequence 1, Application US/08804198  
; Patent No. 5945320  
; GENERAL INFORMATION:  
; APPLICANT: Burgett, Stanley G.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rao, Nagaraja R.  
; APPLICANT: Richardson, Mark A.  
; APPLICANT: Rostek, Paul R., Jr.  
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PAUL R. CANTRELL 1138  
; STREET: LILLY CORPORATE CENTER  
; CITY: INDIANAPOLIS  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46285  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: Macintosh 7.0  
; SOFTWARE: Microsoft Word 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,198  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CANTRELL, PAUL R.  
; REGISTRATION NUMBER: 36,470  
; REFERENCE/DOCKET NUMBER: P9113  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 317-276-3885  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 44377 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 350..14002  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 14046..20036  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 20110..31284  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 31329..36071  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 36155..41830  
US-08-804-198-1

Query Match 6.1%; Score 64.8; DB 2; Length 44377;  
Best Local Similarity 46.7%; Pred. No. 0.0013;  
Matches 382; Conservative 0; Mismatches 422; Indels 14; Gaps 5;

QY 222 ACGCGCGCGCGCGCTCCCGTCTGGCCAGAGCGGCGACTCTACCGCCTCACTATCTG 281  
DB 23568 ACGGATGGCGGCGCTCTCGCGCGCGCGGTCCTCGGGAACCGCGCGGCGACTCCCGCGGCC 23627

QY 282 GACCTGGCGTGGCGACAGTGGGCAATACGTGTGCGCGCGCGCAATGCCATAGGCGAG 341

DB 23628 GGCCTGGCGCGCGCGGGGCAACCGTCTCATCGCGGCGACCTCAGACG-----GTGCGCG 23683  
QY 342 GCCTTCGCTGCGTGGGCTCGAGGTGACGCGGAGSCCGGTGGCGGAGCAGCGCGCG 401  
DB 23684 GCCGACTGGTCCGCTCCTCGAGGACGCGCGGACCGCGTCTCCCTCGTCCCGGACCG 23743  
QY 402 CACTTCTCTGCTCGCGCCACCGTCCATCCGCTGGCGGAGGCTCAGAGGCCACCTTCGCG 461  
DB 23744 ACGCCCCCGCACAGGCGCGCG-CCGCGGACTGACCGGCGTCTCCCTCGTCCCGTGGCG 23802  
QY 462 TCCCGCTGGTGGTCTCCCGAGGCGCGCACTGAGTGTGTCGAAGGAGGCGGCGCGCTG 521  
DB 23803 TCGGACGTCAACGACCGCGCGCACTGGCCCGCTGTTCGACGAGCAGCGCGCCACCGCT 23862  
QY 522 GGTGAGCCCGACGCGCGCGCGCTGCGGTGAGGAGCTCGGCGAGGCAAGTGGCGTGGCG 581  
DB 23863 GCCGTGACGCGCGCGCGCTGTGCTGCGCGCTGCGGAGAGCGGACCGCGCGAC 23922  
QY 582 ATTGGGCGCGCGCGCGCGCGAGCGGCGCACTTACGAGTTCGCGCGCGGAGAACCGCTG 641  
DB 23923 ATCGCGCGCGCGCTCGC---CGCCAAGACCAACGCGCGCGCGCGCGCTGCTGCG 23979  
QY 642 GCGCTGCGACGCGCGCGCGCGCTAGTGTGACTCGGACGCGCGGACACGCGCGCAGC 701  
DB 23980 CCGCGCGCGCGCTCGACGCGCTGTGCTGTCTCTCGTCTCGGAGTGTGGGCGCG 24039  
QY 702 CGCGCGCGGACCTCCAGCGCGCGCTCTGTGCGCACTTGCAGCGCGCGCGCGAGGCTATG 761  
DB 24040 GCGCGCGCGCGCTACG 24099  
QY 762 CGCGCGGAGGCGCGCGCGCG---CTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 818  
DB 24100 CG 24159  
QY 819 ACTGAAGG---CAAGCAGCGCGCGCTCAGCTGTGACTGACCGCGCGCGCGCGCGCG 875  
DB 24160 GCCGACGCTGCGGCGGAGTCTCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 24219  
QY 876 ACGGTGTGAGAGGAGCGCGCGCGCGTGTGACCGAGGCGCGCGCGCGCGCGCGCGCG 935  
DB 24220 CAGCGCGTGCAGCG 24279  
QY 936 GACGCGCAGGAGACTTCTGCTCAAGATCTCTTCTCAAGCAGTGGAGCGCGCGCGCG 995  
DB 24280 GACGTGAGTGGAGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 24339  
QY 996 TACACCTGCACGGGCTCCAACTCGTGGGCGCAGACCTA 1033  
DB 24340 GACGACATCCCGGACGTCAGCGGCTGCGCGCGGCCGA 24377

RESULT 6  
US-07-945-283-1/c  
; Sequence 1, Application US/07945283  
; Patent No. 5352596  
; GENERAL INFORMATION:  
; APPLICANT: Cheung, Andrew K.  
; APPLICANT: Wesley, Ronald D.  
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants  
; TITLE OF INVENTION: Involving The EPO and LIT Genes  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis P. Ribando  
; STREET: 1815 No. 5352596th University Street  
; CITY: Peoria  
; STATE: IL USA  
; COUNTRY: USA  
; ZIP: 61604  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/945,283  
;; FILING DATE: 19920911  
;; CLASSIFICATION: 424  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Ribando, Curtis P  
;; REGISTRATION NUMBER: 27976  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 309-685-4011 ext.513  
;; TELEFAX: 309-685-4128  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 8438 base pairs  
;; TYPE: NUCLEIC ACID  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; HYPOTHETICAL: NO  
;; ANTI-SENSE: NO  
;; ORIGINAL SOURCE:  
;; ORGANISM: Pseudorabies virus  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 622..6495  
;; FEATURE:  
;; NAME/KEY: variation  
;; LOCATION: replace(1099, "g")  
;; FEATURE:  
;; NAME/KEY: variation  
;; LOCATION: replace(1267, "t")  
;; FEATURE:  
;; NAME/KEY: variation  
;; LOCATION: replace(1381, "c")  
;; FEATURE:  
;; NAME/KEY: variation  
;; LOCATION: replace(1566, "c")  
;; FEATURE:  
;; NAME/KEY: variation  
;; LOCATION: replace(7010, "g")  
US-07-945-283-1

Query Match 5.9%; Score 62.6; DB 1; Length 8438;  
Best Local Similarity 42.8%; Pred. No. 0.0033;  
Matches 384; Conservative 0; Mismatches 504; Indels 9; Gaps 1;  
QY 182 CACGCCACAGGTGAGCTGGGAGAGGACACGACCGGTGACGCGCGGCGCGCTTCG 241  
DB 4834 CATGCCCTACCGGCGCGCGCGGACCCCGCGCGGAGCGACCGTCCGAGCCCT 4775  
QY 242 TCTGGCCAGGACCGCGACCTCTACCGCTCACTATCTGGACTGGCGCTGGCGGACG 301  
DB 4774 CTGGCGCGCGCTCCGCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGAGCT 4715  
QY 302 TGGGCAATACGTGCGCGCGCGCAATGCCATAGGCGAGCGCTTCGCTGCGTGGCGCT 361  
DB 4714 GCGGACCGCTGCTGCGCGCTCCGCGCGCTCTGAGCGCTGCTGCGCTGCGCGCT 4655  
QY 362 GCAGGTGACCGGAGCGCGCTGCGCGGACGAGCGCGCGCACTTCCTGCTGGCGCCAC 421  
DB 4654 CCGGCTGCGAGGT 4595  
QY 422 GTCCATCGCTGCGCGGCGCTCAGAGCGCACCTTCGCTGCGCGCGCTGGTGGCTCCC 481  
DB 4594 GACCGCGCGCTGCTGCGCGCTCCGCGACGCGATCCCGGGGCGCGCGCGCGCGGCA 4535  
QY 482 GAGCGCGCGAGTGAAGTGGTCCAAAGGACGCGCGCGCGCTGGGTGAGCGCGCGCGCG 541  
DB 4534 GCAGGCGCGGACTCTGTCGCGCTCGTCCCGCGCACGCTGGCGCGCTGGTGGCTACG 4475  
QY 542 CGTGGCGGTGAGGAGCTGCGCGGAGGCAATGTCGCGATTCGCGGCGCGCGCGCGCG 601  
DB 4474 CGTGACGCGGCG 4415

QY 602 CGACGGCGGACATTACAGAGTCCGGCGCGAGAACCCGCTGGGCGGTGCGAGCGCGCGCG 661  
DB 4414 GGCACAACTGGCGCGCGCGCGCTCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGAGCC 4355  
QY 662 GCGCTAGTGGTGAAGTCCGACGCGCGCGGACACGCGCGCGCGCGCGCGCGCGCGCG 721  
DB 4354 GCGCGCGCGCTGCCCCCTCTGGCCCGAGCAGCGCGCGCTCTGCTGCTCCGAGCGCGCG 4295  
QY 722 CCGCTCTTGGCGACCTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 781  
DB 4294 GCG 4235  
QY 782 CTCACCGCGCGACCG 841  
DB 4234 CACCAAGTCCGCTCCAGCACCAAGTCCAGCTCCGCGCGCGCGCGCGCGCGCGCGCT 4175  
QY 842 CAGCTCTACTGTA-----CGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 892  
DB 4174 CTCGCGCTACGCGAGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4115  
QY 893 CGGCGAGCTGGTGAAGCG 952  
DB 4114 GAAGCG 4055  
QY 953 CGTGTCTCAAGATCTCTTCTGCAAGCAGTCGCGCGCGCGCGCGCGCGCGCGCGCG 1012  
DB 4054 CGGCTCGCGCGCTCCG 3995  
QY 1013 CAACCTCGTGGCGCGAGCCTACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1069  
DB 3994 CCGGCGCAAGCGCGCTCCCTCGGACTCGGCGCGCGCGCGCGCGCGCGCGCGCGCT 3938

## RESULT 7

US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 5.9%; Score 62.6; DB 3; Length 4403765;  
Best Local Similarity 47.0%; Pred. No. 0.0032;  
Matches 231; Conservative 0; Mismatches 254; Indels 6; Gaps 1;  
QY 219 GTACGCGCGCGCGCGCGCTTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2059206  
DB 2059265 GTACG 2059206  
QY 279 CTGGACCTGGCG 338  
DB 2059205 GTGGCGCGGTGGCGCGGTGACGCTGTCGCGCGCGCGCGCGCGCGCGCGCGGT 2059146  
QY 339 GAGGCGCTTCGCTGCG 398

Db 2059145 GGTCCGCGGGCGCGTGGGCTGTTCCGGCTCGTGGGCGCGCGCGCGGAA-- 2059088  
QY 399 CCGACACTTCTGCTGCGGCCACGCTCCATCCGGTGGCGAGGGCTCAGAGGCCACTTTC 458  
Db 2059087 ----TCGGATTGGTCGGGAACGGCGGTGCGGGGGGTCCGGCGGGTCCGCTGCTGG 2059032  
QY 459 CGCTGCGCGGTGGGTGCTCCCGAGGCGCGGCTGAGCTGGTCCCAAGAGGACGGGCGCGCGC 518  
Db 2059031 GCGACGCGGTGCGCGGGCGCGGTGGGTGCGGTCCACTACCGCGGTGCGCGGGG 2058972  
QY 519 CTGGGTGAGCCGACGCGCCCGCGTGGCGGTGAGAGCTCGCGAGGCAAGTGGCTG 578  
Db 2058971 GCGGGCGCAACGCGCGCTGCTGTAGGCGCGCGGGCGCGCGCGCGCACTC 2058912  
QY 579 CGCATTCGGGGCGCGCGCGCGCGGCGGACGCGGCGACTTACGAGGTCCGCGCGGAGAACCCG 638  
Db 2058911 GCGGTGGCTACCGGGGTGGGCGCGCGCGGCGGAAACGGCGCACTGCGGGCTGCTG 2058852  
QY 639 CTGGCGGTGCGACGCGCGCGCGCGCTAGTGGTGAATCGGACGCGCGCGGACACGGCC 698  
Db 2058851 TTGGTGGCGCGCGCGCGCGCGCGCGGATTCGGCTTCGGGTGCGGGCGCGCC 2058792  
QY 699 AGCGCGCGCGG 709  
Db 2058791 GGTGGGTGCGG 2058781

## RESULT 8

US-09-046-992-3  
; Sequence 3, Application US/09046992  
; Patent No. 6140666

## GENERAL INFORMATION:

APPLICANT: Lorberboum-Galski, Haya  
APPLICANT: Yarkoni, Shai  
APPLICANT: Ben-Yehudah, Ahmi  
TITLE OF INVENTION: METHODS OF CANCER DIAGNOSIS  
TITLE OF INVENTION: USING A CHIMERIC TOXIN  
NUMBER OF SEQUENCES: 7

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds, LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY

## COUNTRY: USA

## ZIP: 10036-2811

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/046,992

## FILING DATE: 24-MAR-1998

## CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

## APPLICATION NUMBER:

## FILING DATE:

## ATTORNEY/AGENT INFORMATION:

NAME: Poissant, Brian M  
REGISTRATION NUMBER: 28,462  
REFERENCE/DOCKET NUMBER: 9457-0013-999  
TELEPHONE: 650-493-4935  
TELEFAX: 650-493-5556

## TELEX: 66141 PENNIE

## INFORMATION FOR SEQ ID NO: 3:

## SEQUENCE CHARACTERISTICS:

## LENGTH: 1191 base pairs

## TYPE: nucleic acid

## STRANDEDNESS: single

## TOPOLOGY: linear

## FEATURE:

## NAME/KEY: Coding Sequence

LOCATION: 1...1188  
OTHER INFORMATION:  
US-09-046-992-3

Query Match 5.8%; Score 61.8; DB 3; Length 1191;  
Best Local Similarity 48.0%; Pred. No. 0.0046;  
Matches 207; Conservative 0; Mismatches 222; Indels 2; Gaps 1;

QY 434 GCGCAGGGCTCAGAGGCGACCTTCGGCTGCGGTGGTGGCTCCCGAGGCGCGGAGT 493  
Db 220 GTGACGCGCTGGTGGCTTACTTGGCGGGGCTGTCTGGAACCAAGTCCGACG 279  
QY 494 GAGTGTGTCGAAGACGGGCGCGCTGGGTGAGCC--CGACGGCCCCCGGCGTGGGTG 551  
Db 280 GTGATCCGCAACGCGCTGGCGAGCCCGCGGACGCGCGGCGGACCTGGGCGAAGCATCCG 339  
QY 552 GAGGAGCTCGCGAGGCAAGTGGCTGCGCATTCGGCGGGCGCGCGCGCGCGCGCG 611  
Db 340 GAGCAGCCGAGAGCGCCCTGTCTGGCCCTGACCTTGGCGCGCGCGGAGAGCGCTTC 399  
QY 612 ACTTACGAGGTCCGCGCGCGAGAACCCGCTGGGCGCTGCCAGCGCGCGCGCGCGCTAGT 671  
Db 400 GTCCGGCAGGCAACGCGCAACGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCTG 459  
QY 672 GTGAGACTCGAGCGCGCGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 731  
Db 460 ACCTGCGCGTGGCGCGCGGTGAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 519  
QY 732 GCGACCTGCGAGCGCGCGCGGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 791  
Db 520 GAGCGCAACTATCCCACTGGCGCGGAGTTCTCGCGAGCGCGCGCGCGCGCTCAGC 579  
QY 792 AGCAGCGGACGCGCACCTGACCGGTGACTGAAGCGAGCGAGCGCGCGCGCTGCTAC 851  
Db 580 ACCCGCGGACGCGAGAACTGAGCGGTGGAGCGGTGTCTCCAGGCGCGCGCGCGCGG 639  
QY 852 GTGACCGCGCA 862  
Db 640 GAGCGCGGCTA 650

## RESULT 9

US-08-484-438-41  
; Sequence 41, Application US/08484438  
; Patent No. 5811098  
; Patent No. 5811098 5780031

## GENERAL INFORMATION:

APPLICANT: Plowman, Gregory D.  
APPLICANT: Culouscou, Jean-Michel  
APPLICANT: Shoyab, Mohammed  
APPLICANT: Siegal, Clay B.  
APPLICANT: Hellstr m, Ingegerd  
APPLICANT: Hellstr m, Karl E.  
TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:

## ADDRESSEE: Pennie &amp; Edmonds

## STREET: 1155 Avenue of the Americas

## CITY: New York

## STATE: New York

## COUNTRY: U.S.A.

## ZIP: 10036-2711

## COMPUTER READABLE FORM:

## MEDIUM TYPE: Floppy disk

## COMPUTER: IBM PC compatible

## OPERATING SYSTEM: PC-DOS/MS-DOS

## SOFTWARE: Patent in Release #1.0, Version #1.25

## CURRENT APPLICATION DATA:

## APPLICATION NUMBER: US/08/484,438

## FILING DATE: 07-JUN-1995

## CLASSIFICATION: 530

## PRIOR APPLICATION DATA:

## APPLICATION NUMBER: 08/323,442



```
; FILING DATE: 14-OCT-1994
; APPLICATION NUMBER: US 08/150,704
; FILING DATE: 10-NOV-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/981,165
; FILING DATE: 24-NOV-1992
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 5624-230
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1389 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1386
; US-08-484-438-41

Query Match          5.8%; Score 61.8; DB 1; Length 1389;
Best Local Similarity 48.0%; Pred. No. 0.0046;
Matches 207; Conservative 0; Mismatches 222; Indels 2; Gaps 1;

Qy 434 GCGCAGGGCTCAGAGGCCACTTTCGCTGCGCGTGGTGGTTCCTCCCGAGCGCGCAGT 493
Db 418 GTGCAGCGGTGGTGGCTCTACTGCGCGCGGCTGCTGTTGGAACACAGGTCGACCA 477
Qy 494 GAGCTGTGTCAGAGCAGCGCGCGCTGGGTGAGCC--CGACGCGCGCGCGCTCCG 551
Db 478 GTGATCCGCAACGCCCTGGCGAGCGCGCGGCGGACCTGGGCGAAGCGATCCGC 537
Qy 552 GAGGAGTCTGGCGAGGCAAGTGCCTGCGATTGCGCGCGCGCGCGCGCGCGCGCG 611
Db 538 GAGCAGCGGAGCAGGCGCGCTGTGCGCTGACCTTGGCGCGCGCGAGAGCGGCTTC 597
Qy 612 ACTTACAGAGTCTGGCGCGAGAACCGCTGGGCGCTGCGAGCGCGCGCGCGCGCTAGT 671
Db 598 GTCCGCGAGGCGACCGGCAACAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCTG 657
Qy 672 GTGAGCTCGGACGCGCGCGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCTCTG 731
Db 658 ACCTGCGCGTGGCGCGCGTGAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 717
Qy 732 GCGCACTTGCAGCGCGCGCGCGAGGCTATGCGCGCGAGGCGCGCGCGCGCGCGCG 791
Db 718 GAGCGCAACTATCCCACTGGCGCGGAGTTCCTCGCGAGCGCGCGCGCGCGCGCTCAG 777
Qy 792 AGCAGCGCAGCGCAGCTGACCGTGAAGGCAAGCAGCGCGCGCGCGCGCGCTAC 851
Db 778 ACCCGCGCAGCGCAGAACTGACAGCGTGGAGCGGCTGTCTCCAGGCGCGCGCGCG 837
Qy 852 GTGACCGCGA 862
Db 838 GAGCGCGGCTA 848

RESULT 10
US-08-463-163-2
; Sequence 2, Application US/08463163
; Patent No. 5696237
; GENERAL INFORMATION:
; APPLICANT: Fitzgerald, David J.
; APPLICANT: Chaudhary, Vijay K.
; APPLICANT: Pastan, Ira H.
```

```
; APPLICANT: Waldmann, Thomas A.
; APPLICANT: Queen, Cary L.
; TITLE OF INVENTION: Recombinant Antibody-Toxin Fusion Protein
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,163
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/227,227
; FILING DATE: 22-JAN-1981
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/911,227
; FILING DATE: 24-SEP-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/341,361
; FILING DATE: 21-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/865,722
; FILING DATE: 08-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen L.
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 015280-12211
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1797 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1797
; OTHER INFORMATION: /product= "Anti-Tac(Fv)-PE40"
; US-08-463-163-2
```

```
Query Match          5.8%; Score 61.8; DB 1; Length 1797;
Best Local Similarity 48.0%; Pred. No. 0.0046;
Matches 207; Conservative 0; Mismatches 222; Indels 2; Gaps 1;

Qy 434 GCGCAGGGCTCAGAGGCCACTTTCGCTGCGCGTGGTGGTTCCTCCCGAGCGCGCAGT 493
Db 829 GTGCAGCGGTGGTGGCTCTACTGCGCGCGGCTGCTGTTGGAACACAGGTCGACCA 888
Qy 494 GAGCTGTGTCAGAGCAGCGCGCGCTGGGTGAGCC--CGACGCGCGCGCGCTCCG 551
Db 889 GTGATCCGCAACGCCCTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGATCCGC 948
Qy 552 GAGGAGTCTGGCGAGGCAAGTGCCTGCGATTGCGCGCGCGCGCGCGCGCGCGCG 611
Db 949 GAGCAGCGGAGCAGGCGCGCTGTGCGCTTGAACCGCGCGCGCGCGCGCGCGCTTC 1008
Qy 612 ACTTACAGGCTCGCGCGCGCGAGAACCGCTGGGCGGTGCGAGCGCGCGCGCGCGCTAGT 671
Db 1009 GTCCGCGAGGCGACCGGCAACGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCTG 1068
Qy 672 GTGAGCTCGGACGCGCGCGGACACAGCGCGCGCGCGCGCGCGCGCGCGCGCTCTC 731
```



REGISTRATION NUMBER: 32,944  
REFERENCE/DOCKET NUMBER: 015280-29810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1839 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1839  
OTHER INFORMATION: /product= "Pseudomonas exotoxin A"  
US-09-297-851-1

Query Match 5.8%; Score 61.8; DB 4; Length 1839;  
Best Local Similarity 48.0%; Pred. No. 0.0046;  
Matches 207; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

QY 434 GCGGAGGGCTCAGAGCCACCTTCGCTGCGGCTGGTGGCTCCCGAGGCGGCGAGT 493  
Db 871 GTGACGGCTGGTGGCCCTTACTGGCGGCGGCTGCTGGGAACCAAGTGGACCG 930  
QY 494 GAGCTGTCCAAGACGGGCGCGCTTGGTGGCTCCCGAGGCGGCGAGT 551  
Db 931 GTGATCGCAACGCCCTGGCCAGCCCGGAGCGGCGGACCTGGGCGAAGCGATCCG 990  
QY 552 GAGAGCTCGGAGGAGGAAGTGGCTGGCAATTGGGCGGCGGCGGCGGAGCGGC 611  
Db 991 GAGCAGCGGAGCAGGCGGCTGTGGCCCTGACCTGGCCCGCGCGGAGCGAGCGCTTC 1050  
QY 612 ACTTACAGGTCGCGCGCGAGAACCCGCTGGGCGCTGCCAGCGCGCGCGCGCTAGTG 671  
Db 1051 GTCCGGAGGCGACGGCGAGCAGCAGGCGCGCGCGCGCGCGAGCTGGTGGAGCTG 1110  
QY 672 GTGACTGGAGCGCGCGGACACGCGCAGCGCGCGCGCGCGCGCGCGCGCTCTG 731  
Db 1111 ACCTGCCCGTGGCGCGCGTGAATGGCGGCGCGCGCGCGCGCGCGCGCGCTG 1170  
QY 732 GCGCACCTGACGCGCGCGGAGGCTATGCGCGCGGAGGCGCGCGCGCGCTACCGCC 791  
Db 1171 GAGCGCAACTCCACTGGCGCGGAGTCTCTCGCGACGCGCGCGCGCTAGCTTACG 1230  
QY 792 AGCACCGGACGCGACCTGACGCTGACTGAAGCGAAGCAGCGCGCGCTAGCTGTAC 851  
Db 1231 ACCCGCGCAGCAGAACTGACGCTGAGCGGCTGCTCCAGGCGCACCGCGCAACTGGAG 1290  
QY 852 GTGACCGCGGA 862  
Db 1291 GAGCGCGGCTA 1301

RESULT 13  
US-08-356-786-15  
Sequence 15, Application US/08356786  
Patent No. 5877305  
GENERAL INFORMATION:  
APPLICANT: Huston, James S.  
APPLICANT: Oppermann, Hermann  
APPLICANT: Houston, L. L.  
APPLICANT: Ring, David B.  
TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer  
TITLE OF INVENTION: Marker  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESS: Edmund R. Pitcher, Testa, Hurwitz, & Thibault  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA

ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,786  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/831,967  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pitcher, Edmund R.  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: CRP-053  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1869 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1869  
OTHER INFORMATION: /note= "product = "74isFv-PE40"  
US-08-356-786-15

Query Match 5.8%; Score 61.8; DB 2; Length 1869;  
Best Local Similarity 48.0%; Pred. No. 0.0046;  
Matches 207; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

QY 434 GCGGAGGGCTCAGAGCCACCTTCGCTGCGGCTGGTGGCTCCCGAGGCGGCGAGT 493  
Db 898 GTGACGGCTGGTGGCCCTTACTGGCGGCGGCGCTGCTGGGAACCAAGTGGACCG 957  
QY 494 GAGCTGTCCAAGACGGGCGCGCTTGGTGGCTCCCGAGGCGGCGAGT 551  
Db 958 GTGATCGCAACGCCCTGGCCAGCCCGGAGCGGCGGCGAGCTGGGCGAAGCGATCCG 1017  
QY 552 GAGGAGCTCGGAGGAGTGGCTGGCAATTCGGGCGGCGGCGCGCGCGCGCGCG 611  
Db 1018 GAGCAGCGGAGCGGCGGCTGTGGCCCTGACCTTGGCGCGCGCGCGCGCGCGCTTC 1077  
QY 612 ACTTACAGGCTCGGCGCGGAGAACCCGCTGGGCGCTGCCAGCGCGCGCGCGCGCTAGTG 671  
Db 1078 GTCCGCGAGGCGACCGGCAACGACGAGGCGCGCGCGCGCGCAACCGCGACGCTGGT 1137  
QY 672 GTGACTCGGAGCGCGGAGCAGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCTTC 731  
Db 1138 ACCTGCCCGTGGCGCGGCTGAATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTC 1197  
QY 732 GCGCACCTGACGCGGCGGCGGAGGCTATGCGGCGGCGGAGGCGGCGGCGGCGGCGGCTTC 791  
Db 1198 GAGCGCAACTTCCACTGGCGGAGTCTCTCGGCGACGCGGCGGCGGCGGCTTCAGC 1257  
QY 792 AGCACCGGACGCGACCTTGCACGCTGACTGAAGCGAAGCAGCGGCGGCTAGCTGTAC 851  
Db 1258 AACCGCGCAGCAGAACTGGAGCTGGAGCGGCTGCTCCAGGCGCGCACCGCAACTGGAG 1317  
QY 852 GTGACCGCGGA 862  
Db 1318 GAGCGCGGCTA 1328

RESULT 14  
US-08-743-637B-20  
Sequence 20, Application US/08743637B

Patent No. 5994066  
GENERAL INFORMATION:  
APPLICANT: BERGERON, Michel G.  
APPLICANT: PICARD, Francois J.  
APPLICANT: OUELLETTE, Marc  
APPLICANT: ROY, Paul H.  
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA  
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND  
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED  
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...  
NUMBER OF SEQUENCES: 273  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: QUARLES & BRADY  
STREET: 411 EAST WISCONSIN AVENUE  
CITY: MILWAUKEE  
STATE: WISCONSIN  
COUNTRY: USA  
ZIP: 53202-4497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/743.637B  
FILING DATE: 04-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/526,840  
FILING DATE: 11-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BAKER, Jean C.  
REGISTRATION NUMBER: 35,433  
REFERENCE/DOCKET NUMBER: 850586.90012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 277-5000  
TELEFAX: (414) 277-5591  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2760 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Pseudomonas aeruginosa  
US-08-743-637B-20

Query Match 5.8%; Score 61.8; DB 2; Length 2760;  
Best Local Similarity 48.0%; Pred. No. 0.0046;  
Matches 207; Conservative 0; Mismatches 222; Indels 2; Gaps 1;  
QY 434 GCGGAGGGCTCAGAGCCACCTTCGCTCCGCGTGGTGGCTCCCGAGCGCGCACT 493  
DB 1691 GTGACGGGTGCTGCGCCCTTACCTGGCGCGGCTGTGTGGAACAGTCTGACAC 1750  
QY 494 GAGTGTGTCAAGAGACGGGCGGCGCTGTGGTGAGCC--CGACGGCCCCCGCTGCGCGTG 551  
DB 1751 GTGATCCGCAACGCCCTTGGCGAGCCCGCGGCGGACCTGGCGGAAGCATCCGC 1810  
QY 552 GAGAGCTCGCGAGGCAAGTGGCTCGCATTCGGCGCGCGCGCGCGCGACGCGCGC 611  
DB 1811 GAGACGGGAGAGAGGCGGCTGTGGCCCTGACCTGGCGCGCGCGCGAGAGCGGCTTC 1870  
QY 612 ACTTACAGTTCGCGCGCGAGAACCCGCTGGCGCTGCCAGCGCGCGCGCGCTAGTG 671  
DB 1871 GTCCGGCAGGCGCACCGGCAACGACGAGCGCGCGCGCGCAACCGCGACGTGTGAGCTG 1930  
QY 672 GTGACTCGAGCGCGCGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTG 731  
DB 1931 ACTTGGCCCGTCCGCGCGGTGAATGCGGGGCGCGCGCGGAGCAGCGGCGCGCGCTGTG 1990  
QY 732 GCGCACTGCGAGCGCGCGCGGAGGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 791

DB 1991 GAGCGCAACTATCCCACTGGCGGAGTTCCTCGCGGAGCGCGGCGAGCTCAGCTTCAGC 2050  
QY 792 AGCAGCGGCGGCGGACCTGACGCTGACTGAAGCAAGCAGCGCGCGCTCAGCTGTAC 851  
DB 2051 ACCCGCGGCGGCAAGCAACTGAGCGGTGAGCGGTGCTCCAGGCGGCAACCGCAACTGGAG 2110  
QY 852 GTGACCGGCGA 862  
DB 2111 GAGCGGCGCTA 2121  
RESULT 15  
US-08-526-840B-20  
Sequence 20, Application US/08526840B  
Patent No. 6001564  
GENERAL INFORMATION:  
APPLICANT: BERGERON, Michel G.  
APPLICANT: OUELLETTE, Marc  
APPLICANT: ROY, Paul H.  
TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND  
TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY  
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES  
TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...  
NUMBER OF SEQUENCES: 177  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: QUARLES & BRADY  
STREET: 411 East Wisconsin Avenue  
CITY: Milwaukee  
STATE: Wisconsin  
COUNTRY: USA  
ZIP: 53202-4497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/526,840B  
FILING DATE: 11-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/304,732  
FILING DATE: 12-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BAKER, Jean C.  
REGISTRATION NUMBER: 35,433  
REFERENCE/DOCKET NUMBER: 850586.90012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 277-5000  
TELEFAX: (414) 277-5591  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2760 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Pseudomonas aeruginosa  
US-08-526-840B-20  
Query Match 5.8%; Score 61.8; DB 3; Length 2760;  
Best Local Similarity 48.0%; Pred. No. 0.0046;  
Matches 207; Conservative 0; Mismatches 222; Indels 2; Gaps 1;  
QY 434 GCGGAGGGCTCAGAGCCACCTTCGCTCCGCGTGGTGGCTCCCGAGCGCGCACT 493  
DB 1691 GTGACGGGTGCTGCGCCCTTACCTGGCGCGGCGGCTGTGTGGAACAGTCTGACAC 1750  
QY 494 GAGTGTGTCAAGAGACGGGCGGCGCTGTGGTGAGCC--CGACGGCCCCCGCTGCGCGTG 551  
DB 1751 GTGATCCGCAACGCCCTTGGCGAGCCCGCGGCGGACCTGGCGGAAGCATCCGC 1810



**This Page Blank (uspto)**







QY 852 GTGACCGCGAGCCAGCCAGACCGGTGTGGAAGAGGACGCCAGCTGGTGACCGAG 911  
Db 781 GTGACCGCGAGCCAGCCAGACCGGTGTGGAAGAGGACGCCAGCTGGTGACCGAG 840  
QY 912 GCGCGCGCCACGTGGTGTACGAGGACCGCAGGAGAACTTCGTCTCAAGATCCTCTTC 971  
Db 841 GCGCGCGCCACGTGGTGTACGAGGACCGCAGGAGAACTTCGTCTCAAGATCCTCTTC 900  
QY 972 TGAAGCAGTCCGACCGCGGCTCTACACCTGACCGCGCTCCAACTCTGTGGGCCAGACC 1031  
Db 901 TGAAGCAGTCCGACCGCGGCTCTACACCTGACCGCGCTCCAACTCTGTGGGCCAGACC 960  
QY 1032 TACAGCTCTGTCTGTGTGTAGTCGCGGAGCCCGCGTT 1070  
Db 961 TACAGCTCTGTGTGTGTAGTCGCGGAGCCCGCGTT 999

RESULT 3  
US-10-094-1001/c  
; Sequence 1001, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEXI, NAACHIO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOVUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; PRIOR FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328391  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1001  
; LENGTH: 2884  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-094-749-1001

Query Match 23.8%; Score 255; DB 16; Length 2884;  
Best Local Similarity 100.0%; Pred. No. 1.6e-49;  
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 805 GCACCTGCACGGTGTCTGAAGCAGGACCGCGGCTCAGCTGCTGACCGCGGAGC 864  
Db 634 GCACCTGCACGGTGTCTGAAGCAGGACCGCGGCTCAGCTGCTGACCGCGGAGC 575  
QY 865 CCAAGCCCGAGACGGTGTGGAAGAGGACCGCAGCTGGTACCGAGGCGCGCCACG 924  
Db 574 CCAAGCCCGAGACGGTGTGGAAGAGGACCGCAGCTGGTACCGAGGCGCGCCACG 515  
QY 925 TGGTGTACGAGACCGCAGGAGAACTTCGTCTCAAGATCCTCTTTCGAAGCAGTCGG 984  
Db 514 TGGTGTACGAGACCGCAGGAGAACTTCGTCTCAAGATCCTCTTTCGAAGCAGTCGG 455

QY 985 ACCGCGGCTCTACACCTGACCGGCTCCAACTCTGTGGCCAGACCTACAGCTCTGTGC 1044  
Db 454 ACCGCGGCTCTACACCTGACCGGCTCCAACTCTGTGGCCAGACCTACAGCTCTGTGC 395  
QY 1045 TGGTGTGTAGTCGCG 1059  
Db 394 TGGTGTGTAGTCGCG 380

RESULT 4  
US-09-764-891-6043  
; Sequence 6043, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6043  
; LENGTH: 9591  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-6043

Query Match 18.5%; Score 197.6; DB 10; Length 9591;  
Best Local Similarity 51.7%; Pred. No. 2e-36;  
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

QY 7 CCAGCAGCCACACTCCGGCCCTGCCCCAGAGCCCCCATAGAGAGAGTCCCGCGCCA 66  
Db 2 CCTCCGACCCCCCTGCCCCCACCCTGCGCGTGCAGGCGGTGCGCGCGCGCA 61  
QY 67 CCGTCATGGATCAGCCACAGTTCAGCGGGGCGCCCGCTTTCTCACCGCGCCAGGCT 126  
Db 62 TGAAGCGCAGCTCGGGGATCAGGGGAGCCCCCTGCTTCTCGCGCTTCCCGCGCCTG 121  
QY 127 TCGTGTGTGCTGGCAGGACGCCCTCAGCTGCCAGATCGTGGTAAATCCACGC 186  
Db 122 TCGGGTGTAAAGTGGCGCCGAGGCGAGCTCAAGTGTGTCTTGGGGAGCGCCGC 181  
QY 187 CACAGGTGAGCTGGGAGAGGACCGCAGTCCGCTGACCGCGCGCGCTTCCTGCTGG 246  
Db 182 CTGTAGTGTGTGGGAGAGGCGGCGCAGCTGGCGCTCGGAACGCTGAGCTTC 241  
QY 247 CCCAGCAGCGACCTCTACCGCTCACTATCTGTGACCTGGCGTGGCGGACAGTGGC 306  
Db 242 CGCGCAGCGCGGAGCAGCGCTGCTGCTGACCGCCGCACTGCCACCGCGGGGG 301  
QY 307 AATACGTGTCCGCGCGCAATGCCATAGCGAGGCTTCGCTGCGGTGGCTGCAGG 366  
Db 302 TCTACGTGTCCGCGCGCAACCGCGCGCGGAGGCTTACGCGGCGCGCGCTCACCG 361  
QY 367 TGAACCGGAGCGCGCTGCGCGGAGCGCGGCACTTCTGTGCGGCCACGTTCA 426  
Db 362 TGTGTGAGCGCGCGCTCCGACCCCGAGCTCAGCGCGCGCGCGCTGTCATCGC 421  
QY 427 TCCGCGTGGCGCGAGGCTCAGAGGCCACTTCGCTGCGG-----CGTGGGTGGCTCC 480  
Db 422 CGGGTCCGGAGGCGCGCGCTTCTCAGGGGCTCGATCCAGTGGGTGCTGC 481  
QY 481 CGAGGCGCGAGTGTGTCTCAAGGACGGCGCGCTGGGTGAGCCGAGCGGCCCC 540  
Db 482 GGGGGCGGAGTGTGTGTGAGTGTGCGGGCGGGGCTCCCGAGCCACACTGTACT 541  
QY 541 GCGTGGCGTGGAGGAGCTCGGCGAGGCAAGTGCCTGCGCATTCGGGCGCGCGCCG 600  
Db 542 GGGAGAGGACGGGATGGCCCTTGAGCAAGTGTGGACAGCAGCCACTTCGCGCTCCAG 601  
QY 601 GCGACGCGCGCACTTACGAGTCCGCGCGAGAACCCGCTGGGC--GCTCCAGCGCCG 658

Db 602 CGGGCGCGCGAGGACGGCCCGCGCGAGCGCTGGCACTGCGCATCTTGGCGCTCGGC 661  
QY 659 CGCGCGCTAGTG-GTGAGCTCGAGCGCGCGGAGACGGCCAGCGCGCGCGGACCTCCA 717  
Db 662 TCGCGGATTCGGCGCTCTACGTGTGCCACGCCCGCCCAACGGCGCACGGCGCAGCGG 721  
QY 718 CGCGCGCGCTCTCGCGCACTGTGAGCGCGCGCGCGAGGCTATCGCGCGCGAGCGCGCCC 777  
Db 722 GGGCGCTGCTCAGGTGACACGAGCCCCCGGAGAGCGCCCGCGGACCCGACAGGCCC 781  
QY 778 CGCGCTC-----ACGCGCCAGCACCGGCAACGGCACTGTGAGTGAAGCA 828  
Db 782 CGCGCGCGGTGTGGAGCGCTCAAGTGGCGCGCTTAAGACCTTCTGGGTGAACGAGGCA 841  
QY 829 AGCAGCGCGCTCTCAGCTGTACGTGACCGCGCGAGCCCAAGCGCGAGACGCTGTGGAAGA 888  
Db 842 AGCAGCGCAAGTTCGCTGTCTACGTGTGAGCGCGCGAGCCCGGAGATCGATGGCACT 901  
QY 889 AGCAGCGCACTGTGTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 948  
Db 902 GGGAGGGCGCGCGCTGTCTCCGAGACCGCGCGCGCTCATGTACCGGACCGCGACGGCG 961  
QY 949 ACTTCGTGCTCAAGATCTCTTTCGCAAGCAGTGGAGCGCGCGCGCGCGCGCGCGCG 1008  
Db 962 GCTTCGTGCTCAAGTGTCTTACTGCGCAGGCGCAAGGATCGTGGGCTCTACGTCTGCGCG 1021  
QY 1009 CGTCAACCTCGTGGCGCAGACTACAGCTGTGCTGGTGTGAGTGGCGGAGCGCGCGG 1068  
Db 1022 CGCGCACTCGCGCGCGCAGCGCTCAGTCCGCTGCACTCCAGGTGAAGGTACGCGG 1081

RESULT 5  
US-10-091-438-268  
; Sequence 268, Application US/10091438  
; Publication No. US20030077606A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT217C1  
; CURRENT APPLICATION NUMBER: US/10/091.438  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,879  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/225,757  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/226,868  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/216,647  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/225,267  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/216,880  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/225,270  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/251,869  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/235,834  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/234,274  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: 60/234,223  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: 60/228,924  
; PRIOR FILING DATE: 2000-08-30  
; PRIOR APPLICATION NUMBER: 60/224,518  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,369  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/224,519  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,964  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/241,809  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/249,299  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/236,327  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/241,785  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/244,617  
; PRIOR FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 60/225,268  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,368  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/251,856  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/251,868  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/229,344  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/234,997  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: 60/229,343  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,345  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,287  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,513  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/231,413  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/229,509  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/236,367  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/237,039  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,038  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/236,370  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/236,802  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,037  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,040  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/240,960  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/239,935  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/239,937  
; PRIOR FILING DATE: 2000-10-13

; PRIOR APPLICATION NUMBER: 60/241,787  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/246,474  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/246,532  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/249,216  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,210  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/226,681  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,759  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/225,213  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/227,182  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,214  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/235,836  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/230,438  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/215,135  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 60/225,266  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/249,218  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,208  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,213  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,212  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,207  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,245  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,244  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,217  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,211  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,215  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,264  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,214  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,297  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/232,400  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/231,242  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/232,081  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/232,080  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/231,414  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/231,244  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/233,064  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/233,063  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,397  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,399

; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,401  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/241,808  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,826  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,786  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,221  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/246,475  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/231,243  
; PRIOR FILING DATE: 2000-09-08  
  
Query Match 18.5%; Score 197.6; DB 15; Length 9591;  
Best Local Similarity 51.7%; Pred. No. 2e-36; Indels 19; Gaps 4;  
Matches 558; Conservative 0; Mismatches 504;  
  
QY 7 CCAGCAGCCACACTCCGGCCGCTGCCAGAGCCCCCATTAGAGAGGTCCCGCGGCCA 66  
Db 2 CCCTCGCACCCCCCTTGCCCCCACCCTTGCGCGCTGCAGCGGTGCGCGCGCGCA 61  
  
QY 67 CCGTANTGATCAGCCACAGTTTCAGGGGGGCGCGCTTTCTCACCAGGCAAGGCT 126  
Db 62 TGAAGCGGAGCTCGGGGATCAGGAGAGCCCCCTGCTTCTCGCTTCCCGCGCTG 121  
  
QY 127 TCGTGTGTGCGTGGGCAAGGACGCCCTCAGCTGCCAGATCGTGGTAAATCCACGC 186  
Db 122 TCGGGTGTAGTGGCGGAGGCGAGCTCAAGTGGTGTCTCTGGGGAGCGCGC 181  
  
QY 187 CACAGTGAGCTGGGAGAGGACACAGAGCCGGTGAAGCGCGCGCGCTTCCTGCTGG 246  
Db 182 CTGTAGTGTGGGAGAGGCGGCGAGCTGCGGCTCGGAACCCCTGAGCTTCC 241  
  
QY 247 CCCAGGACGCGACCTTACCGCTCAGCTATCTGACCTGCGCTGGGCGACAGTGGC 306  
Db 242 CGCGGACGCGCGGAGGAGCGGCTGCTGCTGACCGCCCATCTGCCACGAGCGGGG 301  
  
QY 307 AATACGTGTCCGCGCGCAATGCAATAGGCGAGGCTTCTGCTGGGCTGAGG 366  
Db 302 TCTACGTGTCCGCGCGCAACGCGGCGGAGGCTTACGCGGCGCGCGCTCACCG 361  
  
QY 367 TGAACGCGAGGCGCGGTGCGGAGAGCGCGGCTTCTGCTGGGCGGCCAAGTCA 426  
Db 362 TGTGAGCGCGCGGCTTCCGACCGCGGAGCTGAGCGCGCGCGCTGCGCATCGC 421  
  
QY 427 TCCGCGTGGCGAGGCTCAGAGGCGACCTTCCGCTGCGG-----CGTGGGTGGCTCC 480  
Db 422 CGGGGTCCGGGAGGCGCGCGCTTCTCAGCGGGCTCGATCCAGTGGGTGCTGC 481  
  
QY 481 CGAGCGCGGAGTGAAGTGTGCAAGGACGGCGCGGCTTGGGTGAGCGCGCGCGCC 540  
Db 482 GGGGCGGAGGTGTGTGCTGAGCTGCGGGCGGGGGCTTCCCGAGCGCCACACTGTACT 541  
  
QY 541 GGTGCGCTGGAGGAGCTCGCGAGGCAAGTGGCTGGCGATTGGGGGCGCGCGCGC 600  
Db 542 GGGAGAGGACGGGATGCGCTTGGACAGAGTGGGACAGCAGCAGCTTCCGCTCCAGC 601  
  
QY 601 GCGACGCGGCACTTACAGAGTTCGCGCGAGAACCGCTGGGC--GTTGCCAGGCGCGC 658  
Db 602 CGGGCGCGCGGAGGACGCGCGCGGCTGCGACTGCGCATCTTGGGGGTTCGCGC 661  
  
QY 659 CGCGGCGTAGTG--GTGGACTCGGAGCGGAGACAGCGGCGCGCGCGCGCGCGCTCCA 717  
Db 662 TGCCGATTCGGGCTGTACGTGTGACGCGCGGAGCGCGGCGCGCGCGCGCGCGCG 721  
  
QY 718 CGGCGCGCTTCTTGGCGCACCTGACGCGGCGGCGGAGGCTATGCGCGCGCGCGCGCG 777  
Db 722 GGGCGCTGTCTCAGGTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 781  
  
QY 778 CCGCTC-----ACGCGCGAGCACCGGCGACCTGCGCGCTGACTGAGGCA 828



;; PRIOR APPLICATION NUMBER: 60/225,447  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/218,290  
;; PRIOR FILING DATE: 2000-07-14  
;; PRIOR APPLICATION NUMBER: 60/225,757  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/226,868  
;; PRIOR FILING DATE: 2000-08-22  
;; PRIOR APPLICATION NUMBER: 60/216,647  
;; PRIOR FILING DATE: 2000-07-07  
;; PRIOR APPLICATION NUMBER: 60/225,267  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/216,880  
;; PRIOR FILING DATE: 2000-07-07  
;; PRIOR APPLICATION NUMBER: 60/225,270  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/251,869  
;; PRIOR FILING DATE: 2000-12-08  
;; PRIOR APPLICATION NUMBER: 60/235,834  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/234,274  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: 60/234,223  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: 60/228,924  
;; PRIOR FILING DATE: 2000-08-30  
;; PRIOR APPLICATION NUMBER: 60/224,518  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/236,369  
;; PRIOR FILING DATE: 2000-09-29  
;; PRIOR APPLICATION NUMBER: 60/224,519  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/220,964  
;; PRIOR FILING DATE: 2000-07-26  
;; PRIOR APPLICATION NUMBER: 60/241,809  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/249,299  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/236,327  
;; PRIOR FILING DATE: 2000-09-29  
;; PRIOR APPLICATION NUMBER: 60/241,785  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/244,617  
;; PRIOR FILING DATE: 2000-11-01  
;; PRIOR APPLICATION NUMBER: 60/225,268  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/236,368  
;; PRIOR FILING DATE: 2000-09-29  
;; PRIOR APPLICATION NUMBER: 60/251,856  
;; PRIOR FILING DATE: 2000-12-08  
;; PRIOR APPLICATION NUMBER: 60/251,868  
;; PRIOR FILING DATE: 2000-12-08  
;; PRIOR APPLICATION NUMBER: 60/229,344  
;; PRIOR FILING DATE: 2000-09-01  
;; PRIOR APPLICATION NUMBER: 60/234,997  
;; PRIOR FILING DATE: 2000-09-25  
;; PRIOR APPLICATION NUMBER: 60/229,343  
;; PRIOR FILING DATE: 2000-09-01  
;; PRIOR APPLICATION NUMBER: 60/229,345  
;; PRIOR FILING DATE: 2000-09-01  
;; PRIOR APPLICATION NUMBER: 60/229,287  
;; PRIOR FILING DATE: 2000-09-01  
;; PRIOR APPLICATION NUMBER: 60/229,513  
;; PRIOR FILING DATE: 2000-09-05  
;; PRIOR APPLICATION NUMBER: 60/231,413  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/229,509  
;; PRIOR FILING DATE: 2000-09-05  
;; PRIOR APPLICATION NUMBER: 60/236,367  
;; PRIOR FILING DATE: 2000-09-29  
;; PRIOR APPLICATION NUMBER: 60/237,039  
;; PRIOR FILING DATE: 2000-10-02  
;; PRIOR APPLICATION NUMBER: 60/237,038

;; PRIOR FILING DATE: 2000-10-02  
;; PRIOR APPLICATION NUMBER: 60/236,370  
;; PRIOR FILING DATE: 2000-09-29  
;; PRIOR APPLICATION NUMBER: 60/236,802  
;; PRIOR FILING DATE: 2000-10-02  
;; PRIOR APPLICATION NUMBER: 60/237,037  
;; PRIOR FILING DATE: 2000-10-02  
;; PRIOR APPLICATION NUMBER: 60/237,040  
;; PRIOR FILING DATE: 2000-10-02  
;; PRIOR APPLICATION NUMBER: 60/240,960  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/239,935  
;; PRIOR FILING DATE: 2000-10-13  
;; PRIOR APPLICATION NUMBER: 60/239,937  
;; PRIOR FILING DATE: 2000-10-13  
;; PRIOR APPLICATION NUMBER: 60/241,787  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/246,474  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/246,532  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/249,216  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,210  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/226,681  
;; PRIOR FILING DATE: 2000-08-22  
;; PRIOR APPLICATION NUMBER: 60/225,759  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/225,213  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/227,182  
;; PRIOR FILING DATE: 2000-08-22  
;; PRIOR APPLICATION NUMBER: 60/225,214  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/235,836  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/230,438  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR APPLICATION NUMBER: 60/215,135  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: 60/225,266  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/249,218  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,208  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,213  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,212  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,207  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,245  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,244  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,217  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,211  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,215  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,264  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,214  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,297  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/232,400  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/231,242  
;; PRIOR FILING DATE: 2000-09-08

```
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match      18.5%; Score 197.6; DB 15; Length 12415;
Best Local Similarity 51.7%; Pred. No. 1.9e-36;
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

Qy 7 CCAGCAGCCACACTCCGGCGGTGCCCCAGAGCCCCCATAGAGAGAGTCCCGCGCGCA 66
Db 2 CCCTCCGACCCCCCTGCCCCCCCCACCGTTCCGCGCTGCAGCGCGTGGCGCGCGCA 61
Qy 67 CCGTCATGGATCAGCCACAGTTCAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 126
Db 62 TGAAGGCGAGCTCGGGGGATCAGGGGAGCCCCCGCTTCCTCGCGCTTCCCGGGCGTG 121
Qy 127 TCGTGGTGTGGTGGGCAAGGAGCCACCCCTCAGCTGCCAGATCGTGGGTAAATCCACGC 186
Db 122 TCGGGGTGGTAAAGTGGCGCGAGCCGAGCTCAAGTGCCTGCTGGGGGAGCCCGCGC 181
Qy 187 CACAGGTGAGCTGGGAGAGGACCCAGCGCGGTGAGCGCGCGCGCGCTTCGCTTGG 246
Db 182 CTGTAGTGGTGGGAGAGGGGGGCGAGCTGGCGGCTTCGAAAGGTACGCGCG 1081

; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match      18.5%; Score 197.6; DB 15; Length 12415;
Best Local Similarity 51.7%; Pred. No. 1.9e-36;
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

Qy 7 CCAGCAGCCACACTCCGGCGGTGCCCCAGAGCCCCCATAGAGAGAGTCCCGCGCGCA 66
Db 2 CCCTCCGACCCCCCTGCCCCCCCCACCGTTCCGCGCTGCAGCGCGTGGCGCGCGCA 61
Qy 67 CCGTCATGGATCAGCCACAGTTCAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 126
Db 62 TGAAGGCGAGCTCGGGGGATCAGGGGAGCCCCCGCTTCCTCGCGCTTCCCGGGCGTG 121
Qy 127 TCGTGGTGTGGTGGGCAAGGAGCCACCCCTCAGCTGCCAGATCGTGGGTAAATCCACGC 186
Db 122 TCGGGGTGGTAAAGTGGCGCGAGCCGAGCTCAAGTGCCTGCTGGGGGAGCCCGCGC 181
Qy 187 CACAGGTGAGCTGGGAGAGGACCCAGCGCGGTGAGCGCGCGCGCTTCGCTTGG 246
Db 182 CTGTAGTGGTGGGAGAGGGGGGCGAGCTGGCGGCTTCGAAAGGTACGCGCG 1081

; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match      18.5%; Score 197.6; DB 15; Length 12415;
Best Local Similarity 51.7%; Pred. No. 1.9e-36;
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

Qy 7 CCAGCAGCCACACTCCGGCGGTGCCCCAGAGCCCCCATAGAGAGAGTCCCGCGCGCA 66
Db 2 CCCTCCGACCCCCCTGCCCCCCCCACCGTTCCGCGCTGCAGCGCGTGGCGCGCGCA 61
Qy 67 CCGTCATGGATCAGCCACAGTTCAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 126
Db 62 TGAAGGCGAGCTCGGGGGATCAGGGGAGCCCCCGCTTCCTCGCGCTTCCCGGGCGTG 121
Qy 127 TCGTGGTGTGGTGGGCAAGGAGCCACCCCTCAGCTGCCAGATCGTGGGTAAATCCACGC 186
Db 122 TCGGGGTGGTAAAGTGGCGCGAGCCGAGCTCAAGTGCCTGCTGGGGGAGCCCGCGC 181
Qy 187 CACAGGTGAGCTGGGAGAGGACCCAGCGCGGTGAGCGCGCGCGCTTCGCTTGG 246
Db 182 CTGTAGTGGTGGGAGAGGGGGGCGAGCTGGCGGCTTCGAAAGGTACGCGCG 1081

; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match      18.5%; Score 197.6; DB 15; Length 12415;
Best Local Similarity 51.7%; Pred. No. 1.9e-36;
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

Qy 7 CCAGCAGCCACACTCCGGCGGTGCCCCAGAGCCCCCATAGAGAGAGTCCCGCGCGCA 66
Db 2 CCCTCCGACCCCCCTGCCCCCCCCACCGTTCCGCGCTGCAGCGCGTGGCGCGCGCA 61
Qy 67 CCGTCATGGATCAGCCACAGTTCAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 126
Db 62 TGAAGGCGAGCTCGGGGGATCAGGGGAGCCCCCGCTTCCTCGCGCTTCCCGGGCGTG 121
Qy 127 TCGTGGTGTGGTGGGCAAGGAGCCACCCCTCAGCTGCCAGATCGTGGGTAAATCCACGC 186
Db 122 TCGGGGTGGTAAAGTGGCGCGAGCCGAGCTCAAGTGCCTGCTGGGGGAGCCCGCGC 181
Qy 187 CACAGGTGAGCTGGGAGAGGACCCAGCGCGGTGAGCGCGCGCGCTTCGCTTGG 246
Db 182 CTGTAGTGGTGGGAGAGGGGGGCGAGCTGGCGGCTTCGAAAGGTACGCGCG 1081
```

## RESULT 8

```
US-09-764-891-6045
; Sequence 6045, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6045
; LENGTH: 20565
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6045
```

Query Match 18.5%; Score 197.6; DB 10; Length 20565;

Best Local Similarity 51.7%; Pred. No. 1.8e-36;

Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

```
Qy 7 CCAGCAGCCACACTCCGGCGGTGCCCCAGAGCCCCCATAGAGAGAGTCCCGCGCGCA 66
Db 2 CCCTCCGACCCCCCTGCCCCCCCCACCGTTCCGCGCTGCAGCGCGTGGCGCGCGCA 61
Qy 67 CCGTCATGGATCAGCCACAGTTCAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 126
Db 62 TGAAGGCGAGCTCGGGGGATCAGGGGAGCCCCCGCTTCCTCGCGCTTCCCGGGCGTG 121
Qy 127 TCGTGGTGTGGTGGGCAAGGAGCCACCCCTCAGCTGCCAGATCGTGGGTAAATCCACGC 186
Db 122 TCGGGGTGGTAAAGTGGCGCGAGCCGAGCTCAAGTGCCTGCTGGGGGAGCCCGCGC 181
Qy 187 CACAGGTGAGCTGGGAGAGGACCCAGCGCGGTGAGCGCGCGCGCTTCGCTTGG 246
Db 182 CTGTAGTGGTGGGAGAGGGGGGCGAGCTGGCGGCTTCGAAAGGTACGCGCG 241
```

QY 247 CCCAGGCGGACCTTACCGCTCACTATCTTGGACCTTGGCGTGGGCGACAGTGGG 306  
Db |||||  
QY 242 CGGCGGACGGCGGAGACCGCTTGTGTGTACCGCGGACCTGCCACCGACGCGGGG 301  
Db |||||  
QY 307 AATACGTGTGCGCGCGCGCAATGCCATAGCGGAGGCTTTCGTGCGGTGGGCTGACG 366  
Db |||||  
QY 302 TCTACGTGTGCGCGCGCGCAACGCGCGCGGAGGCTTACGCGGCGGCGCGCTCACCG 361  
Db |||||  
QY 367 TGGACGCGGAGGCGCGTGTGCGCGGAGGCGGCGGCTTCTCTGCTGCGGCGGCGGCTCA 426  
Db |||||  
QY 362 TGCTGGAGCGCGCGGCTTCCGACCGCGAGTGTGAGCGCGCGGCGGCGGCTGCTATCGC 421  
Db |||||  
QY 427 TCCCGGTGCGGAGGCTTCAAGGCGGCGGCTTCCGCTGCGG-----CGTGGGTGGTCCC 480  
Db |||||  
QY 422 CGGCGTCCGCGGAGCGCGCGCTTCTTCTTCAAGCGGCGGCTGCGGCGGCGGCGGCTGCTG 481  
Db |||||  
QY 481 CGAGCGCGGAGTGTGCTGCTCAAGGAGCGGCGGCGGCGGCTGCTGCTGAGCGGCGGCGG 540  
Db |||||  
QY 482 GGGGCGGAGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 541  
Db |||||  
QY 541 CGGTGCGGCTGAGGAGTGTGCGGAGGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
Db |||||  
QY 542 GGGAGAGGAGCGGATGCGCTTGAAGAGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGG 601  
Db |||||  
QY 601 CGAGCGGCGGCACTTACGAGTGTGCGCGCGGAGAACCGGCTGCGG-----GCTGCGAGCGCGG 658  
Db |||||  
QY 602 CGGCGCGGCGGAGGAGCGGCGGCGGCGGAGGCTTGGAGTGTGCGGCGGCGGCGGCTGCGG 661  
Db |||||  
QY 659 CGCGCGGCTAGTG-GTGAGTGTGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 717  
Db |||||  
QY 662 TGCGGAGTTCGGGCTGTACGTGTGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 721  
Db |||||  
QY 718 CGGCGCGGCTGCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 777  
Db |||||  
QY 722 GGGCGCTCTCCAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 781  
Db |||||  
QY 778 CGGCGCTC-----ACGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 828  
Db |||||  
QY 782 CGGCGCGGCTGTGAGGCGGCTCAAGTGTGCGGCGGCTTCTGCGGAGGAGGAGGAGGAGG 841  
Db |||||  
QY 829 AGCAGCGGCGCTCAGTGTGTACGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 888  
Db |||||  
QY 842 AGCAGCGGAGTTCGCTGTGTACGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 901  
Db |||||  
QY 889 AGGAGCGGAGTGTGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 948  
Db |||||  
QY 902 GGGAGGCGGCGGCTGTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 961  
Db |||||  
QY 949 ACTTGTGTCAAGATCTTCTTCAAGCAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1008  
Db |||||  
QY 962 GTTCTGTGTCAAGTGTGTCTTCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1021  
Db |||||  
QY 1009 CGTCAAGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1068  
Db |||||  
QY 1022 CGCGCAACTCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1081  
Db |||||

RESULT 9  
US-10-091-438-270/c  
; Sequence 270, Application US/10091438  
; Publication No. US20030077606A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen, et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PTZ17C1  
; CURRENT APPLICATION NUMBER: US/10/091,438  
; CURRENT FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 09/764,879  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628

; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/225,757  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/226,868  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/216,647  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/225,267  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/216,880  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/225,270  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/251,869  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/235,834  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/234,274  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: 60/234,223  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: 60/228,924  
; PRIOR FILING DATE: 2000-08-30  
; PRIOR APPLICATION NUMBER: 60/224,518  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,369  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/224,519  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,964  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/241,809  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/249,299  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/236,327  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/241,785  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/244,617  
; PRIOR FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 60/225,268  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,368  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/251,856  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/251,868  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/229,344  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/234,997  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: 60/229,343  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,345  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,287

```

; PRIOR APPLICATION NUMBER: 60/229,513
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/231,413
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/229,509
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/236,367
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/237,039
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,038
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/236,370
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/236,802
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,037
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215

; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match      18.5%; Score 197.6; DB 15; Length 20565;
Best Local Similarity 51.7%; Pred. No. 1.ee-36;
Matches 558; Conservative 0; Mismatches 504; Indels 18; Gaps 4;

QY 7 CCAGAGCCCACTCCGGCGGTGCCCCAGAGCCCCCATAGAGAGGTCCCCGCGCCCA 66
Db 20564 CCCTCCGACCCCGCCCTGCCCCCACCGTTCCGCCGTGAGGCGGTCCGCGCGCG 20505

QY 67 CCGTCATGATCAGCCACAGTTCAGCGGGGCGCCCGCTTCTCACCCGCCCAAGGCCT 126
Db 20504 TGAAGGCGAGCTCGGGGATCAGGGAGCCCCCGTCTTCTCGGCTTCCCGGGCTG 20445

QY 127 TCGTGGTGTGGTGGCAAGGACCCACCTCAGCTGCCAGATCGTGGGTAATCCACGC 186
Db 20444 TCCGGGTGTAAGTGGCGCGGAGCGGCGAGCTCAAGTSCGTGCTCTGGGAGCGCGCG 20385

QY 187 CACAGGTGAGCTGGGAGAAGACCAAGCGCGGTGAGCGCGCGCGCTTCCGTCTGG 246
Db 20384 CTGTAGTGGTGGGAGAAGGCGGCGAGCTGGCGGCTTCGGAACGCTGAGCTTCC 20325

QY 247 CCCAGGACGGCGACCTCTACCGCTCTACTATCTTGACCTTGGCGCTGGGCGACAGTGGGC 306
Db 20324 CGGCGGACGGCGGAGCACGGCTGTGTGACCGCGCACTGCCACCGACGCGGGG 20265

QY 307 AATACGTGTGCGCGCGCAATGCCATAGCGAGGCTTCGTGCGTGGGCGCTGAGG 366
Db 20264 TCTACGTGTGCGCGCGCAACGCGCGGTACGCGGCGGCGCGCTCACCG 20205

QY 367 TGGACGCGGAGGCGGTGCGCGGAGCAGCGCGCGCTTCTCTGCTGCGCGCCACGCTCCA 426
Db 20204 TGCTGAGCGCGCGGCTCCGACCCCGAGCTGAGCGCGCGCGCTTCCATCGC 20145
```





QY	437	CGAGGGCTCAGAGGCCACCTTCCGCTGCCGCTGGGTGGCTCCCGAGGCGGGCAGCTGAG	496
Db	117	GCGGGGGCGAGGGTGGTCTGACGTGCCGGGCGGGGGGCTTCCCGAGGCCACACTGTA	176
QY	497	CTGGTCCAGAGACGGGGCGGCTTGGGTAGCCGACGGGCCCGCCGCTGCGCTGGAGGA	556
Db	177	CTGGAGAAAGACCGGATGGCCCTTGACCAAGTGTGGACAGCAGCCACTTCGGCGCTCCA	236
QY	557	GCTCGGCGAGGCAAGTGGCTGCGCATTCGGCGCGCGCGCGCGCGACGCGGCACCTTA	616
Db	237	GCCGGGCC-----GCGCCAGAGNACGCCCGCGCGAGCTGGCATTCGTCG	288
QY	617	CGAGGTCGCGGCGAGAAACCCGCTGGGCGCTGCCAGCGCCGCGGGCGGTAGTGTGGA	676
Db	289	GCGGCTCGGCTGCGCGGATTCGCGCGCTACTGTTGCCACGCCGCCAACGCG-----	339
QY	677	CTCGGACGCGCGGGAACACGGCCAGCGCGCGCGGACCTCCACGGCGCGTCTTGGCGCA	736
Db	340	-----CACGGCACGCGAGCGGGGGCGTCTCCAGGTGCACACGCCCCCGAG	390
QY	737	CCTGCAGCGCGGCGGAGGTATGCGCGCCGAGGGCGCCCCCGCTCACGCCCGACAC	796
Db	391	AGCCCGCCGCGAACCCGACAGGCCCC-----GCGCGGTGTGGAGCGGTCAAGTG	446
QY	797	CGGCAACGCGCACTGCACGGTGACTGAAGGCAAGCACGCGCGCTCACTGCTACGTGAC	856
Db	447	CGCGCTTAAGACTCTTGSGGTGAACGAGGGCAAGCACGCCAAGTTCGCTGCTACGTGAT	506
QY	857	GCGGAGGCCAAGCCCGAGACGGTGTGAAGAAGGACGGCCAGCTGTGTACCGAGGCGCG	916
Db	507	GGGCAAGCCGAGCCGAGATCGAATGCACTGGAGGGCGCCCGTGTCTCCCGACCG	566
QY	917	GCGCCACTGTTGTACGAGGACGCGACGAGAACTTCGTCTC	959
Db	567	GCGCGGCTCATGTATCCGCGACCGCGACGGCGGGTGTGTTGCTC	609

## RESULT 12

```

US-10-425-114-26383
; Sequence 26383, Application US/10425114
; Publication No., US20040034888A1
GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 26383
; LENGTH: 1431
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4171-034-C11_FLI
US-10-425-114-26383

```

Db	109	GAAGCAGCTCCGGCGGTGCGACGCTCGGGCCGGGGCGCGTCTGGGGCCCGCTGTGTGGCT	168
Qy	140	GGGCAAGGACGCCACCTCAGCTGCGCAGATCGTGGGTAATCCACAGCCACAGCTGAGCTG	199
Db	169	GGGCTCCGACGAGGCTCGGGCAGCTGTGGCGTCAAGTCTGGCGCGCGCCCGCGGGC	228
Qy	200	GGAGAGGACAGCAGCGGTGACGGCGGGCGCGCTTCGTTGSCCTGSCCAGACACGGCA	259
Db	229	CGCGCGCAGCTGCGACGCGAGGCGCGTCTCCGGGGCCTCTCGTGGCGCACATCT	288
Qy	260	CCTCTACCCGCTCACTATCTCTGGACCTTGGCGCTGGGCGACAGTGGGCAATACTGTGCGC	319
Db	289	GCCCTGCTCGGCTCCCGCGCGGGCGGGCGGAGTACCACTCTGTGGAGTTGCG	348
Qy	320	CGCGCGCAATGCATAGCGGAGGCTTCGTCTGCGTGGGCGCTCGAGTGGAGCGCGAGCG	379
Db	349	GCGCGGGGTGCTGCGCGCAGAGCGCCAGAGCGGGGGGGCGGCTCTCGCGAGCG	408
Qy	380	CGGTGCGCGAGCAGCGCCGCACTTCTGTCTGCGGCCACGTCATCCGCTGCGCGA	439
Db	409	GCCCATCGCGCTACGCGGGGACGTGGGCGCGGGCTGGCTACTCCAGGCGCGTC	468
Qy	440	GGGCTCAGAGGCACTTTCGCTCCG-----CGTGGGTGGCTCCCCGAGGCGCGC	490
Db	469	GCTGTGCACGGGACGTCAAGGCCCGGACGTGGTCACTCGGCGCGCAGGGCGGCGCAG	528
Qy	491	AGTGAGCTGTCTCAAGAGCGGGCGGCGCTGTGGTGAGCCGACGCGCCCGCGTGGCGGT	550
Db	529	GCTGACCGACTTCGTGGGTGCGGAGCGCGCGCGCGGTCTCGACGCGCCCGTCTGGGGCAC	588
Qy	551	GGAGGAGTCTCGCGAGGCAAGTGGCTTGGCCATTCGGGGCGGCGCGCCGCGCGACGGGG	610
Db	589	CCCGCGTTCATGCGCCCGAGGTGGCGCGCGCCAGAGCAGGCCCCCGCGCCGACGT	648
Qy	611	CACTTACGAGTCTCGCGCCGAGAACCCGCTGGCGGTGCGAGCGCGCGCGCGGCGCTAGT	670
Db	649	CTGGCGCTCTGGGTGCATGTGCTGAGCTGGCCACGCGGCGCGGCCCATGAGAGCAGCT	708
Qy	671	GGTGACTCGACGCGCGGACACGGCACCGCGCGCGCGGACCTCCAGCGGCGCGCTCCT	730
Db	709	GGACGAGGACGACCTCTCGCGCGCTCCACCGGATCGGGTACAGGACAGCTGCCGGA	768
Qy	731	GGCGCACCTGACGGCGGGCGGAGGCTATGCGCGCGAGGGGGCCCCCGCTCACGCC	790
Db	769	GGTGCCCGGTGGCTGTGCGCGGAGGCAAGGACTTCTTGGCGCGGTCTTCGAGCGCGC	828
Qy	791	CAGCACCGGCAACGCGCACTTGCAGGTGACTGAAGGCAAGCACGGCGCGCTCAGCTGCTA	850
Db	829	CGCGCGCGCCCGGCCCCACGGCGCGAGCTCGCGGCGCACCCGTTCTGTCGCTCCG	888
Qy	851	CGTGACCGGCGAGCCCAAGCCCCGAGACGTTGTGGAGAGAGGCGCGAGCTGGTGACCGA	910
Db	889	CGCCGCTTGCGCCGCTATCCGGGCCCGGCGAAGCAGGAGGTGGTCCCGTCACCCAAG	948
Qy	911	GGGCGGCGCCACGTGTGTACGAGGACGCGCAGGAGAACTTCGTGCTCAAGAT	964
Db	949	CACGCTGCAACGACGCTTCTGGGACTCTGGACCGCCGAGGACGAAGCGGACGAGAT	1002

RESULT 13

```

RES001 13
US-10-156-761-6483
; Sequence 6483, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHISA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-362

```

```

; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6483
; LENGTH: 2079
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2079)
; US-10-156-761-6483

Query Match      7.4%; Score 79.2; DB 15; Length 2079;
Best Local Similarity 47.7%; Pred. No. 3.9e-09;
Matches 358; Conservative 0; Mismatches 383; Indels 9; Gaps 4;

Qy 65 CACCGTCATGATCAGCCACAGTTCAGCGGGGGCCCGCTTCTCACCGGGCCCAAGGC 124
Db 672 CGCGGGCAGCCCTCGCCCTGCTCTCGCGGGAGCTGGCTGGCGCCCGCCCGCCAGGC 731
Qy 125 CTTCTGCTGTCGCTGGGCAAGGACGCCCTCAGCTGCAGATCGTGGTAAATCCAC 184
Db 732 CGCGCTGCCGACTCCCAAGCAGCTGCGCTCCGAGGGCA---CGAGGGCCCGCGAC 788
Qy 185 GCCACAGGTGAGTGGAGAGAGACAGACCGGTGACGCGCGCGCGCTTCGCTCT 244
Db 789 GGCACGCTCTTCGCGCTGCGCTACGCGGGCTCGCGGCCACCGCCCGAGCGGATCTGCG 848
Qy 245 GCGCCAGGACGGGACCTTACCGCTCAGCTGCAGATCGTGGTAAATCCAC 304
Db 849 CTTCTGCTGCTCGCGCCCGCGGCTGTCGACCGGACACCGCTTCGCGCTCGTGGCG 908
Qy 305 GCAATAGCTGTCGCGCGCGCAATAGCCATAGCGAGCGCTTCGCTCCGCTGGCGCTGCA 364
Db 909 CTGCTCGGTGAGCGGGGCCACACGACCTCGACGATTCGCGCGCTCGCGCTCGTGGCG 968
Qy 365 GGTGACCGGAGGCGGCTGCGGAGCAGGCGCGGCACTTCCTGCTGCGGCGCCAGTGC 424
Db 969 GCTGCTGGAATCCCGCTGCGACAGTACGAGGTGCGCGGCTTCCTGCTGCGCTGCTGGC 1028
Qy 425 CATCCGCTG---CGCGAGGGCTCAGAGGCACTTCCTCGCTGCGGCTGGGTGGCTCCC 480
Db 1029 GCGCTCCCGAGTACGCGAGCGTCCCGCGAGCTTCAGCTGCGCGCGCGCGGATGCT 1088
Qy 481 CGAGCCGCGAGTGTGCTGCTCAAGGACGGGCGCGCTCGGTGAGCGCGCGCGCGCC 540
Db 1089 CGA-GCGGACCGTGGGCTGCTGCACTCTGCGGGCCATCACCAGACGAGCATCG 1147
Qy 541 GCGTGGCGGTGAGGAGCTCGCGAGGCAAGTGGCTGCGATTCGGGCGCGCGCGCGCG 600
Db 1148 AGGCGCGGAGAGTCTCTCGGGATGCGCGGACCGCGCTGTTGGCGCGCGCTCGAGCGGCTG 1207
Qy 601 GCGAGCGGGCACTTACAGAGTCCGCGCGGAGAACCCCGCTGGGGCTGCCAGCGCGCG 660
Db 1208 CCGCGACTGCTGCGCGGTGCGGCGAGCGCGCTGTTGGCGCGCGCTCGAGCGGCTG 1267
Qy 661 CCGCGCTAGTGTGACTCGAGCGCGGAGACAGCGCGCGCGCGGACCTCCAGG 720
Db 1268 CCGAGCGGAGTGGACACCTTGGCGCGCGGCTCATGTCCAGCTGCTGGCGCGATG 1327
Qy 721 CCGCGCTCT-GCGCGACCTGCGAGCGGCGCGGAGGTATGCGCGCGCGAGGGCGCGCC 779
Db 1328 TCGCCACTTCGGCACGAGGCTGCGCGCGGAGCTGTACGGCATCCACCGCTGGTCC 1387
Qy 780 GCTCACCGCGCGAGCACCGGACCGGACCC 809
Db 1388 TCGAGCTGCCGAGCGCGGACCTGCCCC 1417
```

```

RESULT 14
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
; US-10-156-761-1

Query Match      7.4%; Score 79.2; DB 15; Length 9025608;
Best Local Similarity 47.7%; Pred. No. 1.1e-09;
Matches 358; Conservative 0; Mismatches 383; Indels 9; Gaps 4;

Qy 65 CACCGTCATGATCAGCCACAGTTCAGCGGGGGCGCGCTTCTCACCGGGCCCAAGGC 124
Db 7794391 CGCGGGCCAGCCCTCGCCCTGCTCTCGCGGGAGCTGGCTGGCGCGCGCCCGCAGGC 7794450
Qy 125 CTTCTGCTGTCGCTGGGCAAGGACGCCCTCAGCTGCCAGATGTTGGTAAATCCAC 184
Db 7794451 CGCGCTGCGCGACCTCGCCAGCAGCTGGCTCCGAGGGCGA---CGAGGGCCCGCGAC 7794507
Qy 185 GCCACAGGTGAGTGGAGAGGACAGCAGCGGTGAGCGGGCGCGCTTCGCTGCGGCTGCT 244
Db 7794508 GGCACGCTCTTCGCGCTCGCTAGCGCGGGTGGCGGCGCACCGCCAGCGGATCTGCG 7794567
Qy 245 GCGCCAGGACGGGACCTCTACCGCTCACTATCTTGAGACCTGGGCTGGGCGACAGTGG 304
Db 7794568 CTTGCTCTCTCGCGCGCGCGGCTCTGTCGACCGGACACCGCTCGCGCTCGCGCG 7794627
Qy 305 GCAATAGCTGTCGCGCGCGCAATAGCCATAGCGAGGCTTCGCTGCGGCTGGGCTGCA 364
Db 7794628 CTGCTCGGTGAGCGGGCGCCACACGACCTTCGCGCGGCTCGGCTCGTGGCTGCG 7794687
Qy 365 GGTGAGCGGGAGGCGCGCTGCGCGAGCAGCGCGCACTTCCTGCTCGGCGCCACGTC 424
Db 7794688 GCTGCTGGAATCCCCCTCGCCACAGTACGAGTGGCGCGCTCCCTGCTCGCTGCTGCG 7794747
Qy 425 CATCCGCGTG----CGCGAGGGCTCAGAGCGCACCTTCGCTGCGCGCTGGGTGGCTGCC 480
Db 7794748 CGCCCTCGCGAGTACGCGGAGGTCGCCGCGAGCTTCAGCTGGCGCGCGCGGATGCT 7794807
Qy 481 CGAGGCGCGAGTGAAGTGGTCCAGGAGCGGGCGCGCTGGGTGAGCCGCGCGCGCC 540
Db 7794808 CGA-GCGGACCGTGGCGGCTGCTGCTGCTGGCGGCTATCAGGAGCGGACAGATCG 7794866
Qy 541 GCGTGGCGGTGAGGAGCTCGGCGAGGCAAGTGGCTGCGCTGCGCATTCGCGGCGCGCGCG 600
Db 7794867 AGCGCGCGGAGAGCTCTCGGGATGCCCGCGACGCTGCGCTTCGCCGAAACCCCGCGCG 7794926
Qy 601 GCGAGCGCGCACTTACAGGCTCGCGCGCGAGAACCGGCTGGCGGCTGCGAGCGCGCG 660
```



FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (390)..(390)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (392)..(393)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (396)..(417)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (419)..(420)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (422)..(426)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (429)..(430)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (433)..(433)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (435)..(435)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (437)..(438)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (441)..(441)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (447)..(448)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (450)..(450)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (453)..(455)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (460)..(460)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (463)..(466)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (470)..(471)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (474)..(474)  
OTHER INFORMATION: n is a, c, g, or t  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (478)..(480)

Query Match 7.0%; Score 74.6; DB 13; Length 1026;

Best Local Similarity 31.8%; Pred. No. 4.7e-08;  
Matches 200; Conservative 0; Mismatches 429; Indels 0; Gaps 0;  
QY 184 CGCCACAGGTGAGCTGGGAGAAAGACAGACCGGTGACGCCCGCGCGCTTCGGTC 243  
Db 372 CACCNCGCGGGGGGNGNGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNC 431  
QY 244 TGGCCACAGACGGGACCTCTACCGCTCACTATCTGTGACCTGGGCTGGCGACAGTG 303  
Db 432 GNGCNCCNCCCGGNNCCNNNGGCGNCCNNNNNGGNGNCGNNNGNNGNNGGGCG 491  
QY 304 GGCATACTGTGTCGCGCGCGCAATGCCATAGGAGGCTTCCTGCGTGGGCTGC 363  
Db 492 NCNNGNNGGCGNNGGNCNCCNNNGNNGNCCNNNNCGNCCGNNNGGNNNNNAC 551  
QY 364 AGTGGACCGGAGCGCGCTGCGCGGAGACGCGCGCCTCTCTGCTGCGCGCAGT 423  
Db 552 NGNNGNNGGGGCGNNGGNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 611  
QY 424 CCATCGCGTGGCGGAGGCTCAGAGGCGCACCTTCGCTGCGCGCTGGGTGCTCCCGA 483  
Db 612 CGCGNCGCGGNTCCNCCNCCGNNCCNCCNCCNCCNCCNCCNCCNCCNCCGCG 671  
QY 484 GCGCGCAGTGAGTGTCCAAAGGACGGCGCGCGCTGGGTGAGCCCGACGGCGCGCG 543  
Db 672 NGNNGGCGNCGCGNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 731  
QY 544 TGGCGTGGAGAGCTCGCGGAGGCAAGTGCGCTGCGCATTCGGCGCGCGCGCGCG 603  
Db 732 GGGCCNCGGGGNCACGGGGGNGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 791  
QY 604 ACGCGGCACTTACGAGTCCGCGCGGAGAACCGCTGGGCTGCGCGCGCGCGCGCG 663  
Db 792 CCNGCGAAGAGAGCGNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCN 851  
QY 664 CGCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 723  
Db 852 CCNNGCG 911  
QY 724 CGCTCTGGCGCACCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 783  
Db 912 GCGCGNTCACTCG 971  
QY 784 CACCG 812  
Db 972 NNTCAGCG 1000

Search completed: September 20, 2004, 02:51:45  
Job time : 572.829 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 11:16:23 ; Search time 92 Seconds  
(without alignments)  
4471.258 Million cell updates/sec

Title: US-10-077-130-5  
 Perfect score: 41273  
 Sequence: 1 MDOPQSGAPRFTTRPKAFV.....RNRKRRALLYKRHNLAQVR 7968

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB pep.*
5: /cgn2_6/ptodata/2/iaa/PTUTS COMB3 pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1 pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	8423	20.4	1665	4	US-09-858-664A-2	Sequence 2, Appli
2	8423	20.4	1665	4	US-10-274-978-2	Sequence 2, Appli
3	1173	2.8	846	4	US-09-858-664A-3	Sequence 3, Appli
4	1173	2.8	846	4	US-10-274-978-4	Sequence 4, Appli
5	909.5	2.2	549	4	US-09-858-664A-5	Sequence 5, Appli
6	909.5	2.2	549	4	US-10-274-978-6	Sequence 6, Appli
7	892.5	2.2	2860	2	US-08-826-267-2	Sequence 2, Appli
8	892.5	1.4	414	4	US-09-858-664A-13	Sequence 13, Appl
9	592.5	1.4	414	4	US-10-274-978-14	Sequence 14, Appl
10	531	1.3	279	4	US-09-858-664A-4	Sequence 4, Appli
11	531	1.3	279	4	US-10-274-978-5	Sequence 5, Appli
12	521	1.3	298	4	US-09-858-664A-17	Sequence 17, Appl
13	521	1.3	298	4	US-10-274-978-18	Sequence 18, Appl
14	511.5	1.2	508	4	US-09-858-664A-18	Sequence 18, Appl
15	511.5	1.2	508	4	US-10-274-978-19	Sequence 19, Appl
16	504	1.2	250	4	US-09-858-664A-6	Sequence 6, Appli
17	504	1.2	250	4	US-10-274-978-7	Sequence 7, Appli
18	489	1.2	448	2	US-09-159-385-2	Sequence 2, Appli
19	489	1.2	448	3	US-09-186-277-2	Sequence 2, Appli
20	476.5	1.2	11877	3	US-09-105-537-6	Sequence 6, Appli
21	472.5	1.1	260	2	US-07-857-224B-23	Sequence 23, Appl
22	466.5	1.1	454	2	US-09-159-385-1	Sequence 1, Appli
23	466.5	1.1	454	3	US-09-186-277-1	Sequence 1, Appli
24	466	1.1	274	4	US-09-858-664A-14	Sequence 14, Appl
25	466	1.1	274	4	US-10-274-978-15	Sequence 15, Appl
26	465	1.1	358	4	US-09-230-896C-29	Sequence 29, Appl
27	464.5	1.1	1050	4	US-09-428-711A-16	Sequence 16, Appl

## ALIGNMENTS

## RESULT 1

```

US-09-858-664A-2
; Sequence 2, Application US/09858664A
; Patent No. 6482624
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000977-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-858-664A-2

```

	Query Match	20.4%; Score 8423; DB 4; Length 1665;	
	Best Local Similarity	99.9%; Pred.No. 0;	
	Mismatches	0; Mismatches	1; Indels
	Gaps	0;	Gaps
	Matches	1612;	Conservative
Qy	6356	PPSMQVTTIEDVQAQTGTAQFAEIIEGDDPQPSVTWYKDSVOLVDSTRLSQOQGTTYSLV	6415
Db	53	PPSMQVTTIEDVQAQTGTAQFAEIIEGDDPQPSVTWYKDSVOLVDSTRLSQOQGTTYSLV	112
Qy	6416	LHRVASKDAGVYTCLAQNTGGQVLCKAEILLVGCDNEPDSEKSHRRKLHSFVEVKEEIG	6475
Db	113	LHRVASKDAGVYTCLAQNTGGQVLCKAEILLVGCDNEPDSEKSHRRKLHSFVEVKEEIG	172
Qy	6476	RGVGFVKRVQHKGNKILCAAKFIPLRSRTAQAVERDILAALSHPLVTGLLDQPETRK	6535
Db	173	RGVGFVKRVQHKGNKILCAAKFIPLRSRTAQAVERDILAALSHPLVTGLLDQPETRK	232
Qy	6536	TLLIILELCSSELDRLYRKGVTTEAEVKVYIQQLVEGLHYHSGVLHDLIKPSNILM	6595
Db	233	TLLIILELCSSELDRLYRKGVTTEAEVKVYIQQLVEGLHYHSGVLHDLIKPSNILM	292
Qy	6596	VHPAREDIKCDFGPAQNTTPAELOFSQYGSPFEFVSPEIIQQNPVSEASDIWANGVISYL	6655
Db	293	VHPAREDIKCDFGPAQNTTPAELOFSQYGSPFEFVSPEIIQQNPVSEASDIWANGVISYL	352
Qy	6656	SLTCCSPPFAGESDPATILNVLEGRVSWSSPMAAHLSEDADKFITKATLQAPQARPPSAQC	6715
Db	353	SLTCCSPPFAGESDPATIINVLEGRVSWSSPMAAHLSEDADKFITKATLQAPQARPPSAQC	412

QY 6716 LSHPHFLKSMABEAHFTINTKQLFLLARSQWBSLMSYKSIILVMRSIPELLRGPPDPS 6775  
Db 413 LSHPHFLKSMABEAHFTINTKQLFLLARSQWBSLMSYKSIILVMRSIPELLRGPPDPS 472  
QY 6776 LGVAHLCRDTCGGSSSSSSSSSDNELAPPARAKSLPPSPVTHSPLLHPRGFLRPSASLPEE 6835  
Db 473 LGVAHLCRDTCGGSSSSSSSSSDNELAPPARAKSLPPSPVTHSPLLHPRGFLRPSASLPEE 532  
QY 6836 AEASERSTEAPASPEGAGPPAAQCVPHSVIRSLFYHQAGESPEHGALAPGSRHP 6895  
Db 533 AEASERSTEAPASPEGAGPPAAQCVPHSVIRSLFYHQAGESPEHGALAPGSRHP 592  
QY 6896 ARRHLLKGGYIAGALPGLRPLMEHRLVEBEAREEQATLLAKAPFETALRLPASGTH 6955  
Db 593 ARRHLLKGGYIAGALPGLRPLMEHRLVEBEAREEQATLLAKAPFETALRLPASGTH 652  
QY 6956 LAPGSHSLHDSSTPRPSSEACGEAQRPLSPAGPIRDMGHPOGSKQLPSTGGHPG 7015  
Db 653 LAPGSHSLHDSSTPRPSSEACGEAQRPLSPAGPIRDMGHPOGSKQLPSTGGHPG 712  
QY 7016 TAQPERPSDPWQOPAPFCHPKQGSAPQEGCSPHAVAPCPGSPFPGSGCKEAPLPSS 7075  
Db 713 TAQPERPSDPWQOPAPFCHPKQGSAPQEGCSPHAVAPCPGSPFPGSGCKEAPLPSS 772  
QY 7076 PFLGQOPAPPAPAKASPLDCKMGFGDISLPGREKPGFCSPGSGASQASSSSQVSSLRVGS 7135  
Db 773 PFLGQOPAPPAPAKASPLDCKMGFGDISLPGREKPGFCSPGSGASQASSSSQVSSLRVGS 832  
QY 7136 SOVGTEPCPSLDAGWQTQEAEDLSDSTPTLQRPQVTHPKFSLGGRGGYAGVAGTGA 7195  
Db 833 SOVGTEPCPSLDAGWQTQEAEDLSDSTPTLQRPQVTHPKFSLGGRGGYAGVAGTGA 892  
QY 7196 FGGDAGMGLGQPMWARIAMAVSSEEEQBEARAEQSEEQBEARAEPLPOVSARPVP 7255  
Db 893 FGGDAGMGLGQPMWARIAMAVSSEEEQBEARAEQSEEQBEARAEPLPOVSARPVP 952  
QY 7256 EVGAPTRSSPEPTWEDIGQVSLVQIRDLSGDAEADTTSLDISEVDPAAYLNSLDLYI 7315  
Db 953 EVGAPTRSSPEPTWEDIGQVSLVQIRDLSGDAEADTTSLDISEVDPAAYLNSLDLYI 1012  
QY 7316 KYLPFEFMIKPKVKSQAPPPSPMAEELAEPEPTWMPGELGPHAGLEITEESDND 7375  
Db 1013 KYLPFEFMIKPKVKSQAPPPSPMAEELAEPEPTWMPGELGPHAGLEITEESDND 1072  
QY 7376 ALLAEAVGRKRKWSRSLFHPGRLPLDEPAELGLRERKVASVEHISRLKGRPEG 7435  
Db 1073 ALLAEAVGRKRKWSRSLFHPGRLPLDEPAELGLRERKVASVEHISRLKGRPEG 1132  
QY 7436 LEKEGPRKXPGGLASPRLSGLKSWDRAPTLRELSDTVLQSVTLACQVSAQPAQAT 7495  
Db 1133 LEKEGPRKXPGGLASPRLSGLKSWDRAPTLRELSDTVLQSVTLACQVSAQPAQAT 1192  
QY 7496 WSKDGAFLSSSRVLSIATLKNFOLLTLVVAEDLGVTYCSVSNALGTVTTTGVLRKAE 7555  
Db 1193 WSKDGAFLSSSRVLSIATLKNFOLLTLVVAEDLGVTYCSVSNALGTVTTTGVLRKAE 1252  
QY 7556 RPSSSPCPDIEGVYADGVLLVWKPVSYPVTYVQCSLGGGWTTLASDIPCCYLTGK 7615  
Db 1253 RPSSSPCPDIEGVYADGVLLVWKPVSYPVTYVQCSLGGGWTTLASDIPCCYLTGK 1312  
QY 7616 LSRGGTFTTACVSKAGMGFSPSPSVQLLGGPSHLASEEESQGRSAQPLPSTKTFAFQ 7675  
Db 1313 LSRGGTFTTACVSKAGMGFSPSPSVQLLGGPSHLASEEESQGRSAQPLPSTKTFAFQ 1372  
QY 7676 TQQRGFRSVVRCWEKASGRALAAKIIIPHPKDKTAVLREYALXGLRHPHQAUAHY 7735  
Db 1373 TQQRGFRSVVRCWEKASGRALAAKIIIPHPKDKTAVLREYALXGLRHPHQAUAHY 1432  
QY 7736 LSPRHLVILELCSGPELLPCLAEASYSSEVKDYLMQWLSATQYLNHNIHLDRSE 7795  
Db 1433 LSPRHLVILELCSGPELLPCLAEASYSSEVKDYLMQWLSATQYLNHNIHLDRSE 1492  
QY 7796 NMITEYNLLKVDLGNASQSEKVLPSDKFKDYLETWAPELLEGQGAVPQTDIWAIGV 7855

Db 1493 NMITEYNLLKVDLGNASQSEKVLPSDKFKDYLETWAPELLEGQGAVPQTDIWAIGV 1552  
QY 7856 TAFIMLSABEYPVSSSEGCARDLQRLKGLVRLSRVAGLSGGAVALRSTLCAQPMGRPCA 7915  
Db 1553 TAFIMLSABEYPVSSSEGCARDLQRLKGLVRLSRVAGLSGGAVALRSTLCAQPMGRPCA 1612  
QY 7916 SSCLOCPMLTEEGPACSPAPVPTFTARLRFVVRNREKRALLYKRHNLAQVR 7968  
Db 1613 SSCLOCPMLTEEGPACSPAPVPTFTARLRFVVRNREKRALLYKRHNLAQVR 1665

RESULT 2  
US-10-274-978-2  
; Sequence 2, Application US/10274978  
; Patent No. 6670164  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui, et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL000927-CIP-DIV  
; CURRENT APPLICATION NUMBER: US/10/274,978  
; CURRENT FILING DATE: 2002-10-22  
; PRIOR APPLICATION NUMBER: 09/858,664  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 09/711,134  
; PRIOR FILING DATE: 2000-11-14  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 1665  
; TYPE: PRT  
; ORGANISM: Human  
US-10-274-978-2

Query Match 20.4%; Score 8423; DB 4; Length 1665;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1612; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6356 PPSMVTIETDQAOTGGTAQFEALIEGDPQPSVTWYKDSVOLVDSTLSQQCEGTTTSLV 6415  
Db 53 PPSMVTIETDQAOTGGTAQFEALIEGDPQPSVTWYKDSVOLVDSTLSQQCEGTTTSLV 112  
QY 6416 LRHVASKDAGVYTCLAQNTGGQVLCVCAELLVLGGDNEPDEKQSHRRKLHSFYEVKEIG 6475  
Db 113 LRHVASKDAGVYTCLAQNTGGQVLCVCAELLVLGGDNEPDEKQSHRRKLHSFYEVKEIG 172  
QY 6476 RGVFGFVRVQHKNGKILCAAKFIPLSRTRTAQYREDDILAAALSHPIVTGLLDQFETRK 6535  
Db 173 RGVFGFVRVQHKNGKILCAAKFIPLSRTRTAQYREDDILAAALSHPIVTGLLDQFETRK 232  
QY 6536 TLILILELCSSEELDLRLYRKGVVTEAEVKYIQQVLEGLHYLHSHGVHLHDIKPSNLM 6595  
Db 233 TLILILELCSSEELDLRLYRKGVVTEAEVKYIQQVLEGLHYLHSHGVHLHDIKPSNLM 292  
QY 6596 VHPAREDIKICDFGAQNIIPAELOFQYQSPFVSPFETIQONPVSEASDIWANGVLSYL 6655  
Db 293 VHPAREDIKICDFGAQNIIPAELOFQYQSPFVSPFETIQONPVSEASDIWANGVLSYL 352  
QY 6656 SLTCSSPPAGESDRATLNLVLEGRVSNSSPMAHLSEDAKDFIKATLQRAQAPPSAAQC 6715  
Db 353 SLTCSSPPAGESDRATLNLVLEGRVSNSSPMAHLSEDAKDFIKATLQRAQAPPSAAQC 412  
QY 6716 LSHPWFLKSMABEAHFTINTKQLFLLARSQWBSLMSYKSIILVMRSIPELLRGPPDPS 6775  
Db 413 LSHPWFLKSMABEAHFTINTKQLFLLARSQWBSLMSYKSIILVMRSIPELLRGPPDPS 472  
QY 6776 LGVAHLCRDTCGGSSSSSSSSSDNELAPPARAKSLPPSPVTHSPLLHPRGFLRPSASLPEE 6835  
Db 473 LGVAHLCRDTCGGSSSSSSSSSDNELAPPARAKSLPPSPVTHSPLLHPRGFLRPSASLPEE 532  
QY 6836 AEASERSTEAPASPEGAGPPAAQCVPHSVIRSLFYHQAGESPEHGALAPGSRHP 6895



Db 533 AEAERSTEARPASPEGAPPAQGCVPVHRSVIRSLFYHQGESPEHGALAPGSRHP 592  
QY 6896 ARRHLLKGYIAGALPGLREPLMEHRVLEBEAAREEQATLLAKAPSEFETALRLPASGTH 6955  
Db 593 ARRHLLKGYIAGALPGLREPLMEHRVLEBEAAREEQATLLAKAPSEFETALRLPASGTH 652  
QY 6956 LAPGSHSLHDSSTPRPSSAEGEAQRLSPASGGAPIDMGHPGSGKOLSTGGHPG 7015  
Db 653 LAPGSHSLHDSSTPRPSSAEGEAQRLSPASGGAPIDMGHPGSGKOLSTGGHPG 712  
QY 7016 TAQPERPSPDPMQAPFCHPKQGSAPQEGCSPHPAVAPCPGSPFPFGSCKEAPLVPS 7075  
Db 713 TAQPERPSPDPMQAPFCHPKQGSAPQEGCSPHPAVAPCPGSPFPFGSCKEAPLVPS 772  
QY 7076 PFLQOPAPPAPAKASPLDSKMGPGDISLPGRPKPGPCSPGASQASQSSVSLRVS 7135  
Db 773 PFLQOPAPPAPAKASPLDSKMGPGDISLPGRPKPGPCSPGASQASQSSVSLRVS 832  
QY 7136 SOVGTEPSPDADGWTQAEADLSDSTPTLQRPQEQATMRKFSLGGRGVAGVAGYTGA 7195  
Db 833 SOVGTEPSPDADGWTQAEADLSDSTPTLQRPQEQATMRKFSLGGRGVAGVAGYTGA 892  
QY 7196 FGGDAGMLGGPMMWARTAMAVSQSEEEQEAASQSEEQBARAESPLPQVSAAPVP 7255  
Db 893 FGGDAGMLGGPMMWARTAMAVSQSEEEQEAASQSEEQBARAESPLPQVSAAPVP 952  
QY 7256 EVGAPTRSSPEPTWEDIGOVSLVOIRDLSDGAADTISLDISEYDPAVLNLSLDYDI 7315  
Db 953 EVGAPTRSSPEPTWEDIGOVSLVOIRDLSDGAADTISLDISEYDPAVLNLSLDYDI 1012  
QY 7316 KYLPPEFMI FRKVPKSAQPEPPSWABEELAEFFPEPTWMPGELGPHAGLEITESESDVD 7375  
Db 1013 KYLPPEFMI FRKVPKSAQPEPPSWABEELAEFFPEPTWMPGELGPHAGLEITESESDVD 1072  
QY 7376 ALLAEAAVGRKWKSSPSRSLFHPFGRHLPLDPEABELGLRRVKAASVEHISRLIKGRPEG 7435  
Db 1073 ALLAEAAVGRKWKSSPSRSLFHPFGRHLPLDPEABELGLRRVKAASVEHISRLIKGRPEG 1132  
QY 7436 LEKEGPRKKGGLASFRLSGLKSWDRAPTFRLSDETVLQGSVTLACQVSAQPAQAAT 7495  
Db 1133 LEKEGPRKKGGLASFRLSGLKSWDRAPTFRLSDETVLQGSVTLACQVSAQPAQAAT 1192  
QY 7496 WSKGAPLESSRVLISATLKNFOLLITLVVAEDLGVYTCVSNALGTVTTCGLRKA 7555  
Db 1193 WSKGAPLESSRVLISATLKNFOLLITLVVAEDLGVYTCVSNALGTVTTCGLRKA 1252  
QY 7556 RPSSSPCPDIEGVADGVLLVWKPVESYGPVTYIVQCSLEGSWTTLASIDFDCCYLTSK 7615  
Db 1253 RPSSSPCPDIEGVADGVLLVWKPVESYGPVTYIVQCSLEGSWTTLASIDFDCCYLTSK 1312  
QY 7616 LSRGTYTFRACVSKAGMPYSSPSEQVLLGGPSHLASEESQGRSAQPLPSTKTFAFQ 7675  
Db 1313 LSRGTYTFRACVSKAGMPYSSPSEQVLLGGPSHLASEESQGRSAQPLPSTKTFAFQ 1372  
QY 7676 TQIQGRSVVROCKEASGALAAKIIPIHPKDKTAVLREYEAALKGRHPHLAQLHAAY 7735  
Db 1373 TQIQGRSVVROCKEASGALAAKIIPIHPKDKTAVLREYEAALKGRHPHLAQLHAAY 1432  
QY 7736 LSPRLVLILELCSGPELLPCLAEASYSSEVKDYLWQMLSATCYLHNQHILHLDRSE 7795  
Db 1433 LSPRLVLILELCSGPELLPCLAEASYSSEVKDYLWQMLSATCYLHNQHILHLDRSE 1492  
QY 7796 NMITEYNLLKVDIGNAQSLSQBKVLPSDKFKOYLETMAPELLEGGQAVPQTDIWAIGV 7855  
Db 1493 NMITEYNLLKVDIGNAQSLSQBKVLPSDKFKOYLETMAPELLEGGQAVPQTDIWAIGV 1552  
QY 7856 TAFIMLSAEYPVSEGAADLQGRKGLVRLSRVCYAGLSGGAVAFIARSTLCAQPWGRPCA 7915  
Db 1553 TAFIMLSAEYPVSEGAADLQGRKGLVRLSRVCYAGLSGGAVAFIARSTLCAQPWGRPCA 1612  
QY 7916 SSCLOCPWLTBEGPACSRPAPVTFTARLRVVRNREKRALLYKHNLAQVR 7968

Db 1613 SSCLOCPWLTBEGPACSRPAPVTFTARLRVVRNREKRALLYKHNLAQVR 1665  
RESULT 3  
US-09-858-664A-3  
; Sequence 3, Application US/09858664A  
; Patent No. 6482624  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui, et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL000927-CIP  
; CURRENT APPLICATION NUMBER: US/09/858,664A  
; CURRENT FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 09/711,134  
; PRIOR FILING DATE: 2000-11-11  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 846  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-858-664A-3  
Query Match 2.8%; Score 1173; DB 4; Length 846;  
Best Local Similarity 34.9%; Pred. No. 1,1e-54;  
Matches 313; Conservative 121; Mismatches 316; Indels 146; Gaps 29;  
QY 6357 PSMQVTVIIVQAGTGTGAQFAEIIIEGDPQPSVTWKDSVOLVDSTRLSQQQGGTYSLVL 6416  
Db 1 PRFESIMEDVEVGAGETARFAVVVEGKPLPDIMWYKDEVLITSSHSVSENECSLVV 60  
QY 6417 RHVAASDAGVYTCCLAQNTGGVLCXAEILLVGLGDN---EPDSEKQSHR-RKLHSYEVK 6471  
Db 61 LSTGAQGGVYTCQAQNLAGEVSCXAEALAVHSAQTAAMEVGEVGEDEHRCRLSDFYDIH 120  
QY 6472 BEIGRGVGFVKRVQKGNKILCAAKFIPLRSRTAQAYRREDILAAALSHPLVTGLDQF 6531  
Db 121 QEIGRGAFSYLARIVERSGLFEAAKFIPOAKPKASARREARLLARLQCHDCVLYFHEAF 180  
QY 6532 ETRKTLILILELCSSEBELLRLYRKGVVTEAEVKYIQOLVEGLHYLHSHGVHLHDIKPS 6591  
Db 181 ERRRGVLIVTELC-TEELLERJARTVCESEIRAYMROVLEGIHYLHSHGVHLHDVKE 239  
QY 6592 NILMVHFA--REDIKICDFGAQNTIPAELOFSQYSGSPFVSPPEITQQNPVSEASDIAM 6649  
Db 240 NLLVMDGAAGEQVVRICDFGNAQELTPGEPQYQYGTPEFAPEIVNQSPVSGVTDMFV 299  
QY 6650 GVISYLSITCSPFPAGESDRATLLNVLEGRVSWSPMAHLSEDAKDF-IKATLQAPQA 6708  
Db 300 GVVAFLCLTGISFPFVENDRTLLMIRNTYNVAFEETTFLSLSREARGFLIKVLVQ--DRL 357  
QY 6709 RPSAAQCLSHPWFLKSMPEAEAAHFINTKOLKFLARSWRORSIMSYKSLVMRSIPELLR 6768  
Db 358 RPTAETLEHPFKTCAKGAE---VSTDHLKLFISRRRWORSQISYKCHLVLRPPELLR 414  
QY 6769 GPDPSPGLGVARHLCDTTCGSSSSSSSSSSDNEL-----APFARAK-SLPSPSVTH 6816  
Db 415 APPERVWVTMPRR-PPSGGLSSSDSEEEELPSPVRPLQPFSGSRVSLTDIPTED 473  
QY 6817 SPLHPRGFLRSASLPEEAESAESRSTAPAPASPEGAGPAAAGCVPRHSVIR----- 6871  
Db 474 EALGTPETGAATPMDQEQGRAPSOQOEAPSEALLPSPQOEPA--GASPRRELARGSA 532  
QY 6872 -----SLFYHQAG--ESPEHGALAPG-----SRHPRARHLLK 6903  
Db 533 ESALPRAGPRELGRGLHKAASVELPQRRSPGFGATLARGGLGEGEYAQLQALQRLLR 592  
QY 6904 GGIAGALPGLREPLMEH-----RVLEEAAREEQATL-----LAKAPSETALR 6948  
Db 593 GCPEDKVGSLRGPLLESGLGRADPRVARAASAEAPHHPQPLENRGLQKSSSQSGBA 652

```

QY 6949 LPASGTHLAPGHSHLSHDSPTPR-----PSSACGGAQRLPSAPSGGAPIRDMGHPOGS 7004
D5 653 EP-RGRHRRAGAPLEIPVARLGARLQESPSLSALSEAQ--PSSPA-----RPSAP 700
QY 7005 KQLPSTGCHPGCTAOPERPSPDQAPFCHPKQGSAPQEGCSPHPAVAPCPGSPFP 7064
D5 701 K--PST---PKSABFSATTSDAPQPPAP--QPAQDKAPEPRPEPVASKPAP-----PPQ 749
QY 7065 SCK--EAPLVPSPPFLOQ-----PQAPPAPAK-----ASPPLDSKMG 7100
D5 750 ALQTLALPLTFYAQIIQSLQSLGSHAQGPSQGPAPPEPKPHAAVFAVVASPP-----802
QY 7101 GDISLGRPKPGPCSSPGSASQASSQVSSLRVSSOVGTGPGSLDAEGWTQEA 7156
D5 803 -----PGAPEKRVPSAGGPPVLAEKARVPT-----VPPRPGSSLSSTIENLESE 846

RESULT 4
US-10-274-978-4
; Sequence 4, Application US/10274978
; Patent No. 6670164
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP-DIV
; CURRENT APPLICATION NUMBER: US/10/274,978
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 09/858,664
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 846
; TYPE: PRT
; ORGANISM: Human
US-10-274-978-4

Query Match 2.8%; Score 1173; DB 4; Length 846;
Best Local Similarity 34.9%; Pred. No. 1.1e-54;
Matches 313; Conservative 121; Mismatches 316; Indels 146; Gaps 29;

QY 6357 PSMQVTTEDVQAQGGTAQFEAIEGDPQPSVTWYKDSVOLVSTRLSQOQEGTYSVLV 6416
D5 1 PRFSIMEDVEVGAGETARFAVVVEGKPLDIMWKDEVLLTSSHSVFYVENECSLVV 60
QY 6417 RHVASKDAGVYTCLAQNTGGVLCABELLVLGGDN-----EPDSEKQSHR-RKLHSFYEVK 6471
D5 61 LSTGAQDGGVYTCTAQNLAGESVCKAELAVHSAQTAMEVEGVGDEDEHRRRLSDFYDIH 120
QY 6472 BEIGRGVGFVKRYQCHGNKILCAKFIPLRSRTQAQYRERDILAAALSHPLVTGLDDOF 6531
D5 121 QEIGRGAFVLYRRIVERSSGLEPAKFIPOAKPKASAREARLLARLQDCHVLYPHEAF 180
QY 6532 ETRKTLILILELCSBELDLRLYRKGVVTEAEVKYIQQLVEGHYHLHSHVHLIDKPS 6591
D5 181 ERRERGLVIVTELC--TEELLERIAKPTVCESEIRAYMRQVLEGIHYLHSHVHLIDKPS 239
QY 6592 NILMVHQA--REDTKICDFQAQNTIPAELOFSQYSGSPFVSPFIIQQNPVSEASDIWAM 6649
D5 240 NLLVWDGAAGQQVKRICDFNAQELTFCEPQYQYGTGPEFAVEIVNQSPVSGVTILWPV 299
QY 6650 GVISYLSLTCSPPAGESDRATLLNVLEGRVSSPMAHLSBDAXDF--IKATLQAPQA 6708
D5 300 GWAFCLTGISFPVGNDRITLLMIRNYNVAPEETFTLSLSREARGFLIKVLVQ--DRL 357
QY 6709 RPSAAQCLSHWFLKSNPAEAEHINTKQLKFLARSRWQSLMSYKSLIVMSRISIDELL 6768
D5 358 RPTAETLEHPWFKTQAKGAE--VSTDHLKFLSRRRWQSRQISYKCHLVLRPIPELLR 414

```

```

QY 6769 GPPDSFSLGVARHLCDRTGGSSSSSSSSSDNDEL-----APFAEAK-SLPPSPVTH 6816
D5 415 APPERVWVTMPRR--PPSGGLSSSSDSEEELEBLPSVPRFLQPEFGSRVSLTDIPTED 473
QY 6817 SPLLPGRGFLRPSASLPEAEASERSSTEAPAPPASPEGAGPAAQGVPRHSVIR-----6871
D5 474 EALGTFETGAATPMDWQEQGRAPSDQDEAPSPALPSQGEPA--GASPRGELRGSSA 532
QY 6872 -----SLFYHQAG--ESPEHCALAPG-----SRHPARRHLLK 6903
D5 533 ESALPRAGPRELGRGLHKAASVLPORRSPGPGATRLARGGLGEGEYAOQLAQLRQLLR 592
QY 6904 GGYIAGALPGURELMEH-----RVLEEAAREEQATL-----LAKAPFFETALR 6948
D5 593 GCPEDCKVSGURGLPGLLESGLGRARDPRVARAASSEAAHPHPQPLENGLQKSSFSQGEA 652
QY 6949 LPASGTHLAPGHSHLSHDSPTPR-----PSSACGGAQRLPSAPSGGAPIRDMGHPOGS 7004
D5 653 EP-RGRHRRAGAPLEIPVARLGARLQESPSLSALSEAQ--PSSPA-----RPSAP 700
QY 7005 KQLPSTGCHPGCTAOPERPSPDQAPFCHPKQGSAPQEGCSPHPAVAPCPGSPFP 7064
D5 701 K--PST---PKSABFSATTSDAPQPPAP--QPAQDKAPEPRPEPVASKPAP-----PPQ 749
QY 7065 SCK--EAPLVPSPPFLOQ-----PQAPPAPAK-----ASPPLDSKMG 7100
D5 750 ALQTLALPLTFYAQIIQSLQSLGSHAQGPSQGPAPPEPKPHAAVFAVVASPP-----802
QY 7101 GDISLGRPKPGPCSSPGSASQASSQVSSLRVSSOVGTGPGSLDAEGWTQEA 7156
D5 803 -----PGAPEKRVPSAGGPPVLAEKARVPT-----VPPRPGSSLSSTIENLESE 846

RESULT 5
US-09-858-664A-5
; Sequence 5, Application US/09858664A
; Patent No. 6482624
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-664A-5

Query Match 2.2%; Score 909.5; DB 4; Length 549;
Best Local Similarity 38.3%; Pred. No. 8.1e-41;
Matches 204; Conservative 88; Mismatches 181; Indels 59; Gaps 12;

QY 6274 TGTSEAPAVPRVPQPLLHEGFEQEPFAIAQAEWTVPIRMGAAMPAGT-----GE 6326
D5 25 TVKSSSKSPSEPPVQLLEHGFTLE--EAPAMLDKPDIVVVEGQ--PASVTVTENHVEAQ 81
QY 6327 LLVDVHSHVRETTQRTTYQAID-----THTAR-----6355
D5 82 VVWVSCRGALLAEARAGVYELSQPDQDDQYCLRICVSRDRMGALCTCTARNHGTCTCVTL 141
QY 6356 ----PPSMQVTTEDVQAQGGTAQFEAIEGDPQPSVTWYKDSVOLVSTRLSQOQEGT 6411
D5 142 ELAEAPRESIMEDVEVGAGETARFAVVVEGKPLDIMWKDEVLLTSSHSVFYVENE 201
QY 6412 YSLVLRHVASKDAGVYTCLAQNTGGVLCABELLVLGGDN-----EPDSEKQSHR-RKLHS 6466

```

Db 202 CSLVVLSTGADGGVYTCTAQNLAGVSCXAEALVHSAQTAAMEVGVGEDEHRRRLSD 261  
Qy 6467 FYEYKEBIGRQVGVKVRQKGNKILCAAKFIPLKSRTRAQYRREDILAAALSHPLVTG 6526  
Db 262 FYDIHQBIGRGAFAFYLRIVERSGLEGFAAKFIPSOAKPKASARREARLLARLQHDVLY 321  
Qy 6527 LLDOPETRKTLILILELCSSEELLDRLYRGVTVTEAVKVIQOLVGLHSHGVHLH 6586  
Db 322 FHEAFERRRGLVIVTELCTEELLERIAKPTVCESIRAYMQVLEGIHYHQSHVLHL 380  
Qy 6587 DIKPSNIMLVHFA--REDIKICDFGFAQNTPAELQFSQYSGSPFVSPETIQONPVSEAS 6644  
Db 381 DVKENLLVWDGAAGEQQVRICDFGNAQELTPGEPQYCYGTPEFVAPEIVNQSPVSGVT 440  
Qy 6645 DIWAMGVISYLSLTCSSPPAGESDRATLLNVLEGRVSWSSPMAHLSBDKDF-IXATLQ 6703  
Db 441 DIWPGVVAFLCLTGISPFVGENDRITLMNIRYNVAFETTFLSLSREARGFLIKVLVQ 500  
Qy 6704 RAPQARPSAQCLSHPWFLKSMPEAEAHFINTKQKFLARSRWQSRMSYK 6755  
Db 501 --DLRPTAETLEHPWFKTQAKGAE--VSTDHLKFLSRRRWQSRQISYK 547  
RESULT 6  
US-10-274-978-6  
; Sequence 6, Application US/10274978  
; Patent No. 6670164  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui, et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CLC00927-CIP-DIV  
; CURRENT APPLICATION NUMBER: US/10/274,978  
; CURRENT FILING DATE: 2002-10-22  
; PRIOR APPLICATION NUMBER: 09/859,664  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 09/711,134  
; PRIOR FILING DATE: 2000-11-14  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 549  
; TYPE: PRT  
; ORGANISM: Human  
US-10-274-978-6  
Query Match 2.2%; Score 909.5; DB 4; Length 549;  
Best Local Similarity 38.3%; Pred. No. 8.1e-41;  
Matches 204; Conservative 89; Mismatches 181; Indels 59; Gaps 12;  
Qy 6274 TGTSEAPAVPRVPQPLIHGEGEQEPEALARAQEWTPIRMGEAAMPAGCT-----GE 6326  
Db 25 TVKSSSKPSPSEPVLLEHGTLE-EAPALMDKPDIVYVVEGQ--PASVTVTFNHVEAQ 81  
Qy 6327 LLMDVHSHVRETTQRTTYQAID-----THTAR----- 6355  
Db 82 VVMRSRGALLERAGVYELSQPDQDQYLCIRSVRRDMGALTCTARNRHGTQTCVTL 141  
Qy 6356 ----PPSMQVITEDVQAQGTGTAQFEAIIEDGPOPSVTWYKDSVQLVDSTRLSQOQSGTT 6411  
Db 142 ELAARPRETIMEDVEVCAGETAFANVVEGKPLPDINWYKDEVLLTSSHSVSVYENE 201  
Qy 6412 YSLVRHVASKDAGVYTCLANTGGQVLCKAELLVLGDN----EPDSEKQSHR-RKLHS 6466  
Db 202 CSLVVLSTGADGGVYTCTAQNLAGVSCXAEALVHSAQTAAMEVGVGEDEHRRRLSD 261  
Qy 6467 FYEYKEETGRGVGVKVRQKGNKILCAAKFIPLKSRTRAQYRREDILAAALSHPLVTG 6526  
Db 262 FYDIHQEGRGAFAFYLRIVERSGLEGFAAKFIPSOAKPKASARREARLLARLQHDVLY 321  
Qy 6527 LLDOPETRKTLILILELCSSEELLDRLYRGVTVTEAVKVIQOLVGLHSHGVHLH 6586

Db 322 FHEAFERRRGLVIVTELCTEELLERIAKPTVCESIRAYMQVLEGIHYHQSHVLHL 380  
Qy 6587 DIKPSNIMLVHFA--REDIKICDFGFAQNTPAELQFSQYSGSPFVSPETIQONPVSEAS 6644  
Db 381 DVKENLLVWDGAAGEQQVRICDFGNAQELTPGEPQYCYGTPEFVAPEIVNQSPVSGVT 440  
Qy 6645 DIWAMGVISYLSLTCSSPPAGESDRATLLNVLEGRVSWSSPMAHLSBDKDF-IXATLQ 6703  
Db 441 DIWPGVVAFLCLTGISPFVGENDRITLMNIRYNVAFETTFLSLSREARGFLIKVLVQ 500  
Qy 6704 RAPQARPSAQCLSHPWFLKSMPEAEAHFINTKQKFLARSRWQSRMSYK 6755  
Db 501 --DLRPTAETLEHPWFKTQAKGAE--VSTDHLKFLSRRRWQSRQISYK 547  
RESULT 7  
US-08-826-267-2  
; Sequence 2, Application US/08826267  
; Patent No. 5994070  
; GENERAL INFORMATION:  
; APPLICANT: Streuli, Michel  
; TITLE OF INVENTION: No. 5994070e1 TRIO Molecules and Uses Related Thereto  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/826,267  
; FILING DATE: 1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/014,214  
; FILING DATE: 27 MARCH (1996)  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Amy E. Mandragouras  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: DFN-010  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)227-5941  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2860 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-826-267-2  
Query Match 2.2%; Score 892.5; DB 2; Length 2860;  
Best Local Similarity 24.7%; Pred. No. 6.9e-39;  
Matches 318; Conservative 209; Mismatches 493; Indels 265; Gaps 42;  
Qy 5562 DSDSDSKTSPASPRHGRSR----PSSIQESSSESED--GDARGEIDFIVVVTADYLP- 5614  
Db 1758 DSDSDSAATQDSTVEGRNGLSGTISKSSSGMQSCGEEGE-----EGADAVPLP 1811  
Qy 5615 ---GAFQDAITREGQYVEVLDAHPLRLVLTPTKS--SPSRQGWSP-----AY 5661  
Db 1812 PPMAIQHSLLQPDQSQ----DDKASSRLVR--PTSSTPSAAELVSAIBELVSKMAL 1864  
Qy 5662 LDRLKLSFENWGAAPAF-PGE-----AVSEDEYKAEKLS-----VIELLSSEQA 5707  
Db 1865 EDRPSSLVDQDSSSPFNPSDNLSSSPIDEMERKSSSLKRRHYVLQELVETED 1924



```

; PRIOR APPLICATION NUMBER: 09/711,134
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Human
US-10-274-978-14

Query Match
  1.4%; Score 592.5; DB 4; Length 414;
Best Local Similarity 34.5%; Pred. No. 5.6e-24;
Matches 145; Conservative 69; Mismatches 191; Indels 15; Gaps 6;

QY 6356 PPSQVNTIEDVQAGTGAFAIEEDGDPQSVTVWYKDSVOLVD-----STRLSQQOEGT 6410
DB 1 PPEVIFLSEVTCETGTETVVRRCVGRPKASITWKQPEHNTLNNDGHYSISYSDLGAT 60

QY 6411 TYSIVLRHVASKDAGVYTCAQNTGGQVLCKAELLVLGGNEPDSEKSHRRKLHGFYEV 6470
DB 61 --LKIVGVTTEDDGIYTCIAVNDMGSSASSASLRVLGPG--MDGIMVTWKDNFDSFYSE 115

QY 6471 KEEICRGVGFVXVQHKIKLCAAKFIPLRSETRACAYRERDILAAALSHPLVTGLDQ 6530
DB 116 VAEIGRGVFSYVKCDQKGTQVATKPVANKMKRQDVTHELGILQSLQHPHLLGLDT 175

QY 6531 PETRKTILILELCSSEBLLDRYKGVVTEAEVYVYIQQLVEGLHYLHSHGVHLHDKP 6590
DB 176 PETFTSYLVLEMAQDQGLLDCVVRWGLSTEGKIRAHLEGEVLEAVRYLHNCRIAHLDKP 235

QY 6591 SNILMVPH-AREDIKIDCFGAQNTITPAELQFSQYSGPEFVSPEHIOQNPVSEASDIWAM 6649
DB 236 ENILVDESLEAKPTIKLADFGDAVQLNTYYIHQLGNPEFAAPEIILGNPVSILTSQTSV 295

QY 6650 GVISVLSITCSPSPAGSDRATLNVLEGRVSWSSPMAAHLSEDAKDFIKATLQRAQAR 6709
DB 296 GVLTYVLLSGVSPFLDDSVETCLNICELDFSFPDDYFKGVSKAKFVCFLLQEDPAKR 355

QY 6710 PSAAQCLSHPWFLKSPABEAHFINTKQLFLARSQW---RSLMYSKILVNRSTPEL 6766
DB 356 PSAAALAEQW-LQAGNCRSTGVLDTSRLTSFIERKQNDVVRPIRSIKNFLOSLRLPRV 414

RESULT 10
US-09-858-664A-4
; Sequence 4, Application US/09858664A
; Patent No. 6482624
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-664A-4

Query Match
  1.3%; Score 531; DB 4; Length 279;
Best Local Similarity 39.5%; Pred. No. 6.4e-21;
Matches 111; Conservative 45; Mismatches 123; Indels 2; Gaps 1;

QY 7639 SPSEQVLLGPGSHLASEESQGRSAQPLPSTTKTFATQIQGRFSVVRQCKEASGRAL 7698
DB 1 SPAKEVVSPPGSSPRSSPRPEGTTLRQGPQKPYTFLEEKARGFGVVRACRENATGRTF 60

QY 7699 AAKIIPYHPKDKTAVLRREYEAALKGLRHPHLAQHAAVLSPRHLVLILELCSGPELLPCA 7758
DB 61 VAKIVPYAAEGKPRVLQYEYEVLTLLHHERIMSLHEAVITPRVLVLAESCGNRELLCGLS 120

QY 7759 ERASYSESEVKDYLMQMLSATQYLHNQHLHLDLRSENMIITEYNLLKVVLDLGNQAQSLQ 7818
DB 121 DRFRYSEDDVATYVMVQLQGLDYLHGHVHLHLDIKPDNLLAPDNALKIVDFGSAQPNP 180

QY 7819 EKVLPSEDKFDYLETMAPELLEGQGAVPQTDIWAIGVTAFTAFIMLSAEYVPSSEGARDLQRG 7878
DB 181 QALRPLGHRGTGLEFWAPEVWKGEPIGSAITDIWAGVLTIVMLSGRSPFYEPDPQETEAR 240

QY 7879 LRKGLVRLSRCYAGLSGGAVALFRLSTLCQAPWGRPCASSCL 7919
DB 241 IVGGRFADFQLYPNTSQSATLFLRKVLSVHPWSPR--SSCL 279

RESULT 12
US-09-858-664A-17
; Sequence 17, Application US/09858664A
; Patent No. 6482624
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
```

Query Match 1.3%; Score 521; DB 4; Length 298;  
Best Local Similarity 35.2%; Pred. No. 2.4e-20;  
Matches 105; Conservative 68; Mismatches 123; Indels 2; Gaps 2;  
US-09-858-664A-17

6462 RKLHSFVEVKEEIGRGVGFVKVQHKGNKILCAAKFIPLRSRTRAQAYR-ERDILAALS 6520  
1 QKVSDFYDIERLGSFGQVFLVEKTRKWAGKFFKAYSAKEKNIQKISIMNCLH 60

6521 HPLVTGLLDQFETRKTLLILELCSSEELDLRLYRKG-VTEAEVKYIQQLVEGLHYLH 6579  
61 HPKLVQCVDAFEKANIWMVLEIVSGGELFERIIDFELTERECIKYMRQISEGVEYIH 120

6580 SHGVHLHDIKPSNIMLVHPAREDIKIDFGFAQNTIPAELOFSQYSGSPFVSPELIQONP 6639  
121 KQGIHVLDLKPENIMCVNKTGTRIKLIDFGLARRLENAGSLKVLFGTPEFVAPEVINYEP 180

6640 VSEASDIWAMGVI SYLSLTCSPPFAGESDRATLLNVLEGRVSWSSPMAAHLSEDAKDFIK 6699  
181 ISYATDMWSIGVICVILVSGLSPFMGNDNETLANVTSATWDFDDEAFDEISDDAKDFIS 240

6700 ATLQAPQAPSAQAQCLSHPWFLKMPABEAHFNITKOLKFLARSRQORSLMSYKSI 6757  
241 NLLKDKMKNRLDCTCQLOHPMLMDTKMNEAKKSKDKRMKMYMARRKWKQKTGNVRAI 298

US-09-858-664A-17

Query Match 1.3%; Score 521; DB 4; Length 298;  
Best Local Similarity 35.2%; Pred. No. 2.4e-20;  
Matches 105; Conservative 68; Mismatches 123; Indels 2; Gaps 2;  
US-10-274-978-18

6462 RKLHSFVEVKEEIGRGVGFVKVQHKGNKILCAAKFIPLRSRTRAQAYR-ERDILAALS 6520  
1 QKVSDFYDIERLGSFGQVFLVEKTRKWAGKFFKAYSAKEKNIQKISIMNCLH 60

6521 HPLVTGLLDQFETRKTLLILELCSSEELDLRLYRKG-VTEAEVKYIQQLVEGLHYLH 6579

Query Match 1.2%; Score 511.5; DB 4; Length 508;  
Best Local Similarity 25.7%; Pred. No. 1.7e-19;  
Matches 125; Conservative 75; Mismatches 170; Indels 117; Gaps 5;  
US-09-858-664A-18

6354 ARPPSMQVTIEDVQAQTGTAQFEAIEGDPQPSVTWYKDSVQLVDSRLSQQOEGTYS 6413  
21 AMPQIIQFPEDQKVRAGESVELFGKVTGTQPICTWTKFRKQIQDSEHIKVENSENSGK 80

6414 LVLHRVASKDAGVYTCLAQNTGGQVLCABELLVLG----- 6448  
81 LTILAARQEHGCGCYTLLENKLGSRQAQVNLTVVDPKPPAGTFCASDIRSSSLTLSWYG 140

6449 -----CDNE 6452

141 SSYDGSVAQVSYSIEIWDSSANKTWKELATCSTSFNVQDILLPHYEYKFRVAINVGTSE 200

6453 P-----DSEKQ-----SHRRKLSHFVEVKEEIGRGV 6478

201 PSQSELTTVGKEPEEPKMKWRCQTDDEKEPEVDYRTVTINTEQKQDFDYDIERLGSKG 260

6479 FGFKRVQHKGNKILCAAKFIPLRSRTRAQAYR-ERDILAALSPLVTGLLDQFETRKTLL 6537

261 FGQVFLVEKTRKWAGKFFKAYSAKEKNIQKISIMNCLHHPKLVQCVDAFEKANI 320

6538 ILILELCSSEELDLRLYRKG-VTEAEVKYIQQLVEGLHYLHSHGVHLHDIKPSNIMLV 6596

321 VMVLEIVSGGELFERIIDFELTERECIKYMRQISEGVEYIHKQGIHVLDLKPENIMCV 380

6597 HPAREDIKIDFGFAQNTIPAELOFSQYSGSPFVSPELIQONPVSEASDIWAMGVI SYLS 6656

381 NKTGTRIKLIDFGLARRLENAGSLKVLFGTPEFVAPEVINYEFISATDMWSIGVICYL 440

6657 LTCSSPPAGESDRATLLNVLEGRVSWSSPMAAHLSEDAKDFIKATLQAPQAPSAQAQCL 6716

441 VSGLSPPMGNDNETLANVTSATWDFDDEAFDEISDDAKDFISNLLKDKMKNRLDCTCQCL 500

QY 6717 SHPWFLK 6723  
 Db 501 QHPWLMK 507

## RESULT 15

US-10-274-978-19  
 ; Sequence 19, Application US/10274978  
 ; Patent No. 6670164  
 ; GENERAL INFORMATION:  
 ; APPLICANT: WEI, Ming-Hui, et al  
 ; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
 ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
 ; TITLE OF INVENTION: THEREOF  
 ; FILE REFERENCE: CLC00927-CIP-DIV  
 ; CURRENT APPLICATION NUMBER: US/10/274,978  
 ; CURRENT FILING DATE: 2002-10-22  
 ; PRIOR APPLICATION NUMBER: 09/858,664  
 ; PRIOR FILING DATE: 2001-05-17  
 ; PRIOR APPLICATION NUMBER: 09/711,134  
 ; PRIOR FILING DATE: 2000-11-14  
 ; NUMBER OF SEQ ID NOS: 34  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 19  
 ; LENGTH: 508  
 ; TYPE: PRT  
 ; ORGANISM: Human  
 US-10-274-978-19

Query Match 1.2%; Score 511.5; DB 4; Length 508;  
 Best Local Similarity 25.7%; Pred. No. 1.7e-19;  
 Matches 125; Conservative 75; Mismatches 170; Indels 117; Gaps 5;  
 QY 6354 ARPPSMQVIEDVQAOQTGTAQFAIIIEGDPQPSVWYKQSVQLVSTRLSQOQEGTYS 6413  
 Db 21 AMPQIIQFPEDQKVRAGESVELFGKVTGTPICTWKFQIQDSEHIKVENSENGSK 80  
 QY 6414 LVLHVASKAGVYVTCIAQNTGGQVLCKAEILLVG----- 6448  
 Db 81 LTLAARQEHGCGYTLLENKLGSRQAOVNLTVVDKDPDPAGTPCASIIRSSSLTWSYG 140  
 QY 6449 -----GNE 6452  
 Db 141 SSYDGSVAQSVSIEIWDANKTWKELATCRSTSFNVQDLLPDHEYKFRVRAINVYGTSE 200  
 QY 6453 P-----DSEKQ-----SHRRKLHSFYEVKEEIGRGV 6478  
 Db 201 PSQESLTVGKPEEPKQKRCQTDDEKPEVDYRTVTINTEQKVSDFYDIERLGS GK 260  
 QY 6479 FGFVKRVQHGKNIKILCAAKFIPLRSRTRAQAYR-ERDILAALSHPLVTGLLDQFETRTL 6537  
 Db 261 FQVFLVEKTRKWAGKFFKAYSAKEKENIRQEIISIMNCLHHPKLVQCVDAFEERANI 320  
 QY 6538 ILILELCSSEELDLRYKGV-VTEAEVKVYIQQLVEGLHYLHSHGVLDIKFSNLMV 6596  
 Db 321 VMVLEIVSGGELFERIIDDFELTERECIKYMRQISEGVEYHKQGIHVHLDLKPENINCV 380  
 QY 6597 HPAREDIKICDFGAQNTIPAELOFSQXGSPFVSPEIIQQNPVSEASDIWANGVISYLS 6656  
 Db 381 NKTGTRIKIDFGLARLENAGSLKVLGTPEFVAFEVINYEPISYATDMNSIGVICYL 440  
 QY 6657 LTCSSPFPAGESDRATILNVLEGRVSWSPMAHLSDEADFIKATLQAPQARPSAAOCL 6716  
 Db 441 VSGLSFPFGNDNDNETLVNTSATWDFDDEAFDEISDDAKDFISNLLKKDMKNRLDCTOCL 500  
 QY 6717 SHPWFLK 6723  
 Db 501 QHPWLMK 507

Search completed: September 13, 2004, 11:37:13  
 Job time : 98 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 19:28:00 ; Search time 199.275 Seconds  
(without alignments)  
9751.690 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_16862\_17246

Perfect score: 385  
Sequence: 1 ccaggcgagatcttgaca.....acctgcagcactggagcgc 385

Scoring table:

IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	385	100.0	7893	14	US-10-077-130-2
2	385	100.0	7928	15	US-10-307-019-5
3	385	100.0	8106	14	US-10-077-130-1
4	385	100.0	23907	14	US-10-077-130-6
5	385	100.0	24120	14	US-10-077-130-4
6	377	97.9	628	11	US-09-864-408A-1271
7	47.4	12.3	1290	15	US-10-156-761-393
8	47.4	12.3	9025608	15	US-10-156-761-1
9	42.6	11.1	4180	13	US-10-342-887-1266
10	42.6	11.1	4180	13	US-10-172-118-1266
11	42.6	11.1	4180	16	US-10-439-388-10
12	42.6	11.1	4180	17	US-10-776-827-93
13	42.6	11.1	4282	9	US-09-778-927A-16
14	42.6	11.1	4366	14	US-10-044-090-668

Sequence 46, Appl  
Sequence 45, Appl  
Sequence 97552, A  
Sequence 4858, Ap  
Sequence 1, Appli  
Sequence 11431, A  
Sequence 21800, A  
Sequence 21887, A  
Sequence 99, Appl  
Sequence 523, App  
Sequence 26, Appl  
Sequence 1172, Ap  
Sequence 35, Appl  
Sequence 7923, Ap  
Sequence 13198, A  
Sequence 78150, A  
Sequence 31562, A  
Sequence 2574, Ap  
Sequence 97551, A  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 1133, Ap  
Sequence 680, App  
Sequence 3618, Ap  
Sequence 22, Appl  
Sequence 22, Appl  
Sequence 22, Appl  
Sequence 22, Appl  
Sequence 22, Appl  
Sequence 22, Appl

US-09-965-528-46  
US-09-969-984-46  
US-10-437-963-97552  
US-10-156-761-4858  
US-10-156-761-1  
US-10-282-122A-11431  
US-09-918-395-21800  
US-09-918-395-21887  
US-09-764-898-99  
US-10-264-049-523  
US-09-764-898-26  
US-10-104-047-1172  
US-10-203-295-35  
US-10-757-701-7923  
US-10-282-122A-13198  
US-10-437-963-97551  
US-10-260-238-1246  
US-10-425-114-31562  
US-09-867-701-2574  
US-10-437-963-97551  
US-10-159-257A-2  
US-10-252-198-1  
US-10-412-699B-1133  
US-10-374-780A-680  
US-10-156-761-3618  
US-09-909-320-22  
US-09-909-088B-22  
US-09-905-231A-22  
US-09-902-853-22  
US-09-907-824-22  
US-09-907-841-22

#### ALIGNMENTS

#### RESULT 1

US-10-077-130-3  
; Sequence 3, Application US/10077130  
; Publication No. US20020168742A1  
; GENERAL INFORMATION:  
; APPLICANT: Kapeller-Libermann, Rosana  
; APPLICANT: Acton, Susan L.  
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family  
; TITLE OF INVENTION: Members and Uses Therefor  
; FILE REFERENCE: MPI2001-047PIRCPI(M)  
; CURRENT APPLICATION NUMBER: US/10/077,130  
; CURRENT FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: 60/269201  
; PRIOR FILING DATE: 2001-02-15  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 7893  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-10-077-130-3

Query Match 100.0%; Score 385; DB 14; Length 7893;  
Best Local Similarity 100.0%; Pred. No. 5.8e-102;  
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGCGAGATCTTTGACATCTAGTGGTCCCGCTGACTACCTGCCCTAGGGGCTGA 60  
777 CCGAGCGAGATCTTTGACATCTAGTGGTCCCGCTGACTACCTGCCCTAGGGGCTGA 836

QY 61 GGAGGATGCCATCAGCTCGGGAAGCCAGATGTGGAGGTCTCTGGATGACGCCACCC 120  
837 GGAGGATGCCATCAGCTCGGGAAGCCAGATGTGGAGGTCTCTGGATGACGCCACCC 896

QY 121 ACTGCGCTGGTGTTCGCGACCAAGCCCAAGTCCAGCCCTCAGCGAGGTGGGT 180  
897 ACTGCGCTGGTGTTCGCGACCAAGCCCAAGTCCAGCCCTCAGCGAGGTGGGT 956

QY 181 GTCACAGGCTACCTGGACAGAGGCTCAAGCTGTCACTGTGAGTGGGGCCCGCTGAGGC 240  
DB 957 GTCACAGGCTACCTGGACAGAGGCTCAAGCTGTCACTGTGAGTGGGGCCCGCTGAGGC 1016  
QY 241 CCCTGAGTTCCTGGGAGGCTGTGTCTGAAGACGAATACAAGSCAAGGCTGAGCTCTGT 300  
DB 1017 CCCTGAGTTCCTGGGAGGCTGTGTCTGAAGACGAATACAAGSCAAGGCTGAGCTCTGT 1076  
QY 301 GATCCAGAGCTGCTGAGTTCGAGCAGGCTTCGTCGAGGAGCTGAGTTCCTGAGAG 360  
DB 1077 GATCCAGAGCTGCTGAGTTCGAGCAGGCTTCGTCGAGGAGCTGAGTTCCTGAGAG 1136  
QY 361 CCACCACTGCAGCACCTGGAGCGC 385  
DB 1137 CCACCACTGCAGCACCTGGAGCGC 1161

## RESULT 2

US-10-307-019-5  
; Sequence 5, Application US/10307019  
; Publication No. US20030108533A1  
; GENERAL INFORMATION:  
; APPLICANT: Zeng, Wenlin  
; APPLICANT: Stanton, Lawrence  
; APPLICANT: SCIOS, INC.  
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION  
; FILE REFERENCE: SCIOS.021DV1  
; CURRENT APPLICATION NUMBER: US/10/307,019  
; CURRENT FILING DATE: 2002-11-26  
; PRIOR APPLICATION NUMBER: 09/548,473  
; PRIOR FILING DATE: 2000-04-13  
; PRIOR APPLICATION NUMBER: 60/129,552  
; PRIOR FILING DATE: 1999-04-16  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 7928  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (60)...(7847)  
US-10-307-019-5

Query Match 100.0%; Score 385; DB 15; Length 7928;  
Best Local Similarity 100.0%; Pred. No. 5.8e-102;  
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGGCGAGATCTTTGACATCTACGTGTGTCACCGCTGACTACTCTGCCCTAGGGGCTGA 60  
DB 734 CCGAGGCGAGATCTTTGACATCTACGTGTGTCACCGCTGACTACTCTGCCCTAGGGGCTGA 793  
QY 61 GCAGGATGCCATCAGCTGCGGAAAGCCAGTATGTGGAGTCTCTGGATGCAAGCCACCC 120  
DB 794 GCAGGATGCCATCAGCTGCGGAAAGCCAGTATGTGGAGTCTCTGGATGCAAGCCACCC 853  
QY 121 ACTGCGTGTGCTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGCAGGGCTGGGT 180  
DB 854 ACTGCGTGTGCTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGCAGGGCTGGGT 913  
QY 181 GTCACAGGCTACTGTCGACAGGAGGCTCAAGCTGTCACTGAGTGGGGGCGCGCTGAGGC 240  
DB 914 GTCACAGGCTACTGTCGACAGGAGGCTCAAGCTGTCACTGAGTGGGGGCGCGCTGAGGC 973  
QY 241 CCCTGAGTTCCTGGGAGGCTGTGTCTGAAGACGAATACAAGSCAAGGCTGAGCTCTGT 300  
DB 974 CCCTGAGTTCCTGGGAGGCTGTGTCTGAAGACGAATACAAGSCAAGGCTGAGCTCTGT 1033  
QY 301 GATCCAGAGCTGCTGAGTTCGAGCAGGCTTCGTCGAGGAGCTGAGTTCCTGAGAG 360  
DB 1034 GATCCAGAGCTGCTGAGTTCGAGCAGGCTTCGTCGAGGAGCTGAGTTCCTGAGAG 1093

QY 361 CCACCACTGCAGCACCTGGAGCGC 385  
DB 1094 CCACCACTGCAGCACCTGGAGCGC 1118

## RESULT 3

US-10-077-130-1  
; Sequence 1, Application US/10077130  
; Publication No. US20020168742A1  
; GENERAL INFORMATION:  
; APPLICANT: Kapeller-Libermann, Rosana  
; APPLICANT: Acton, Susan L.  
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family  
; FILE REFERENCE: MPI2001-047PIRCP1(W)  
; CURRENT APPLICATION NUMBER: US/10/077,130  
; CURRENT FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: 60/269201  
; PRIOR FILING DATE: 2001-02-15  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 8106  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: (1)...(71)  
; NAME/KEY: CDS  
; LOCATION: (72)...(7964)  
; NAME/KEY: 3'UTR  
; LOCATION: (7965)...(8106)  
US-10-077-130-1

Query Match 100.0%; Score 385; DB 14; Length 8106;  
Best Local Similarity 100.0%; Pred. No. 5.8e-102;  
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGGCGAGATCTTTGACATCTACGTGTGTCACCGCTGACTACTCTGCCCTAGGGGCTGA 60  
DB 848 CCGAGGCGAGATCTTTGACATCTACGTGTGTCACCGCTGACTACTCTGCCCTAGGGGCTGA 907  
QY 61 GCAGGATGCCATCAGCTGCGGAAAGCCAGTATGTGGAGTCTCTGGATGCAAGCCACCC 120  
DB 908 GCAGGATGCCATCAGCTGCGGAAAGCCAGTATGTGGAGTCTCTGGATGCAAGCCACCC 967  
QY 121 ACTGCGTGTGCTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGCAGGGCTGGGT 180  
DB 968 ACTGCGTGTGCTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGCAGGGCTGGGT 1027  
QY 181 GTCACAGGCTACTGTCGACAGGAGGCTCAAGCTGTCACTGAGTGGGGGCGCGCTGAGGC 240  
DB 1028 GTCACAGGCTACTGTCGACAGGAGGCTCAAGCTGTCACTGAGTGGGGGCGCGCTGAGGC 1087  
QY 241 CCCTGAGTTCCTGGGAGGCTGTGTCTGAAGACGAATACAAGSCAAGGCTGAGCTCTGT 300  
DB 1088 CCCTGAGTTCCTGGGAGGCTGTGTCTGAAGACGAATACAAGSCAAGGCTGAGCTCTGT 1147  
QY 301 GATCCAGAGCTGCTGAGTTCGAGCAGGCTTCGTCGAGGAGCTGAGTTCCTGAGAG 360  
DB 1148 GATCCAGAGCTGCTGAGTTCGAGCAGGCTTCGTCGAGGAGCTGAGTTCCTGAGAG 1207  
QY 361 CCACCACTGCAGCACCTGGAGCGC 385  
DB 1208 CCACCACTGCAGCACCTGGAGCGC 1232

## RESULT 4

US-10-077-130-6  
; Sequence 6, Application US/10077130  
; Publication No. US20020168742A1  
; GENERAL INFORMATION:  
; APPLICANT: Kapeller-Libermann, Rosana

```
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047PIRCPI(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 23907
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-077-130-6

Query Match          100.0%; Score 385; DB 14; Length 23907;
Best Local Similarity 100.0%; Pred. No. 6.5e-102;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGGGGAGATCTTTGACATCTACGTGGTCCAGCGCTGACTACCTGCCCCCTAGGGGCTGA 60
Db 16791 CCGAGGGGAGATCTTTGACATCTACGTGGTCCAGCGCTGACTACCTGCCCCCTAGGGGCTGA 16850

QY 61 GCAGATGCCATCAGCTGGCGGGAAGGCCAGTATGTGGAGTCTCTGATGAGCCCAACC 120
Db 16851 GCAGATGCCATCAGCTGGCGGGAAGGCCAGTATGTGGAGTCTCTGATGAGCCCAACC 16910

QY 121 ACTGCGTGGCTGTTCGACACCAAGCCCAAGTCCAGCCCTCAGCGGAGGGCTGGT 180
Db 16911 ACTGCGTGGCTGTTCGACACCAAGCCCAAGTCCAGCCCTCAGCGGAGGGCTGGT 16970

QY 181 GTCACAGCCCTACCTGGACAGAGGCTCAAGCTGTCACTGAGTGGGGGGCGCTGAGGC 240
Db 16971 GTCACAGCCCTACCTGGACAGAGGCTCAAGCTGTCACTGAGTGGGGGGCGCTGAGGC 17030

QY 241 CCCTGAGTTCCTGGGAGGCTGTCTGAGAGCAATACAAAGCAAGGCTGAGCTCTGT 300
Db 17031 CCCTGAGTTCCTGGGAGGCTGTCTGAGAGCAATACAAAGCAAGGCTGAGCTCTGT 17090

QY 301 GATCCAGGAGCTGTGAGTCTTGAGCAGGCGCTTCGTGGAGAGCTGCAGTTCTTGCAGAG 360
Db 17091 GATCCAGGAGCTGTGAGTCTTGAGCAGGCGCTTCGTGGAGAGCTGCAGTTCTTGCAGAG 17150

QY 361 CCACCACTTGACGACCTGGAGCGC 385
Db 17151 CCACCACTTGACGACCTGGAGCGC 17175
```

```
RESULT 5
US-10-077-130-4
; Sequence 4, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MPI2001-047PIRCPI(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 24120
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(71)
; NAME/KEY: CDS
```

```
; LOCATION: (72)...(23978)
; NAME/KEY: 3'UTR
; LOCATION: (23979)...(24120)
US-10-077-130-4

Query Match          100.0%; Score 385; DB 14; Length 24120;
Best Local Similarity 100.0%; Pred. No. 6.5e-102;
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGGGGAGATCTTTGACATCTACGTGGTCCAGCGCTGACTACCTGCCCCCTAGGGGCTGA 60
Db 16862 CCGAGGGGAGATCTTTGACATCTACGTGGTCCAGCGCTGACTACCTGCCCCCTAGGGGCTGA 16921

QY 61 GCAGATGCCATCAGCTGGCGGGAAGGCCAGTATGTGGAGTCTCTGATGAGCCCAACC 120
Db 16922 GCAGATGCCATCAGCTGGCGGGAAGGCCAGTATGTGGAGTCTCTGATGAGCCCAACC 16981

QY 121 ACTGCGTGGCTGTTCGACACCAAGCCCAAGTCCAGCCCTCAGCGGAGGGCTGGT 180
Db 16982 ACTGCGTGGCTGTTCGACACCAAGCCCAAGTCCAGCCCTCAGCGGAGGGCTGGT 17041

QY 181 GTCACAGGCTACCTGGACAGAGGCTCAAGCTGTCACTGAGTGGGGGGCGCTGAGGC 240
Db 17042 GTCACAGGCTACCTGGACAGAGGCTCAAGCTGTCACTGAGTGGGGGGCGCTGAGGC 17101

QY 241 CCCTGAGTTCCTGGGAGGCTGTCTGAGAGCAATACAAAGCAAGGCTGAGCTCTGT 300
Db 17102 CCCTGAGTTCCTGGGAGGCTGTCTGAGAGCAATACAAAGCAAGGCTGAGCTCTGT 17161

QY 301 GATCCAGGAGCTGTGAGTCTTGAGCAGGCGCTTCGTGGAGAGCTGCAGTTCTTGCAGAG 360
Db 17162 GATCCAGGAGCTGTGAGTCTTGAGCAGGCGCTTCGTGGAGAGCTGCAGTTCTTGCAGAG 17221

QY 361 CCACCACTTGACGACCTGGAGCGC 385
Db 17222 CCACCACTTGACGACCTGGAGCGC 17246
```

```
RESULT 6
US-09-864-408A-1271
; Sequence 1271, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Enc
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1271
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(1)
; OTHER INFORMATION: wherein n may be a, c, g or t
US-09-864-408A-1271
```

```
Query Match          97.9%; Score 377; DB 11; Length 628;
Best Local Similarity 100.0%; Pred. No. 9.5e-100;
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 AGATCTTTGACATCTACGTGGTCCAGCGCTGACTACCTGCCCCCTAGGGGCTGAGCAGATG 68
Db 2 AGATCTTTGACATCTACGTGGTCCAGCGCTGACTACCTGCCCCCTAGGGGCTGAGCAGATG 61
QY 69 CCATCAGCTGCGGGAAGGCCAGTATGTGGAGTCTCTGATGAGCCCAACCCTGCGCT 128
```

Db 62 CCATCACGCTCGGGAAGCCAGTATGTGGAGTCTTGATGACGCCACCCACTGGCT 121  
 Qy 129 GGCCTGTCCGCACCAAGCCCAAGTCCAGCCCTCAGCGCAGGCTGGGTGTCACCG 188  
 Db 122 GGCCTGTCCGCACCAAGCCCAAGTCCAGCCCTCAGCGCAGGCTGGGTGTCACCG 181  
 Qy 189 CCTACTGGACAGAGGCTCAAGCTGTCACTGTAGTGGGGGCGCTGAGGCCCTGAGT 248  
 Db 182 CCTACTGGACAGAGGCTCAAGCTGTCACTGTAGTGGGGGCGCTGAGGCCCTGAGT 241  
 Qy 249 TCCCTGGGAGGCTGTGTCTGAAGACGAATACAGGCAAGGCTGAGTCTGTGATCCAGG 308  
 Db 242 TCCCTGGGAGGCTGTGTCTGAAGACGAATACAGGCAAGGCTGAGTCTGTGATCCAGG 301  
 Qy 309 AGCTGCTGAGTTCTGAGCAGCCCTTCGTGGAGGAGTGCAGTTCTCTGACAGGCCACCC 368  
 Db 302 AGCTGCTGAGTTCTGAGCAGCCCTTCGTGGAGGAGTGCAGTTCTCTGACAGGCCACCC 361  
 Qy 369 TGCAGCACTGGAGGC 385  
 Db 362 TGCAGCACTGGAGGC 378

## RESULT 7

US-10-156-761-393  
 ; Sequence 393, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; PRIOR FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 393  
 ; LENGTH: 1290  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(1290)

Query Match 12.3%; Score 47.4; DB 15; Length 1290;  
 Best Local Similarity 49.0%; Pred. No. 0.00084;  
 Matches 126; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
 Qy 18 ACATCTACGTGGTCCCGCTGACTACCTGCCCCCTAGGGGCTGAGCAGGATGCCATCACGC 77  
 Db 305 ACATCGACCTCGACGACGCGACCGACTGCGGTCGGGAGTACGTGCTTACCTCAAGG 364  
 Qy 78 TGGGGAGGCGCAGTATGTGGAGTCTGTGATGACGCCACCCACCTGCGTGGCTGTGTC 137  
 Db 365 TCTGGGAACGCCAGCTGACCGAGTCCAGGATCCAGCCCTCGTCTCCAGAACCCGCGCTCGGC 424  
 Qy 138 GCACCAAGCCCAAGTCCAGGCCCTCAGCGCAGGCTGGGTGTCCACGACCTACTCGG 197  
 Db 425 TGCACCAAGCCCAAGTCCAGGCCCTCAGCGCAGGCTGGGTGTCCACGACCTACTCGG 484  
 Qy 198 ACAGGAGGCTCAAGTGTACCTGTAGTGGGGGCGCGTGAAGCCCTGAGTTCCTTGGGG 257  
 Db 485 TCAAGAGCGGCGAAGAACCCCGTGCACCGGATCCCGCGGAGCAGCAGCCGCGGCGCA 544

Qy 258 AGGCTGTGTCTGAAGAC 274  
 Db 545 CGGGCGTCTCGACGTC 561

## RESULT 8

US-10-156-761-1  
 ; Sequence 1, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; PRIOR FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 1  
 ; LENGTH: 9025608  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (4187715)  
 ; OTHER INFORMATION: a, t, c, g, other or unknown

## US-10-156-761-1

Query Match 12.3%; Score 47.4; DB 15; Length 9025608;  
 Best Local Similarity 49.0%; Pred. No. 0.0016;  
 Matches 126; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
 Qy 18 ACATCTACGTGGTCCCGCTGACTACCTGCCCCCTAGGGGCTGAGCAGGATGCCATCACGC 77  
 Db 471875 ACATCGACCTCGACGACGCGACCGACTGCGGTCGGGAGTACGTGCTTACCTCAAGG 471934  
 Qy 78 TCGGGAGGCGCAGTATGTGGAGTCTGTGATGACGCCACCCACCTGCGTGGCTGTGTC 137  
 Db 471935 TCTGGGAACGCCAGCTGACCGAGTCCAGGATCCAGCCCTCGTCTCCAGAACCCGCGCTCGGC 471994  
 Qy 138 GCACCAAGCCCAAGTCCAGGCCCTCAGCGCAGGCTGGGTGTCCACGACCTACTCGG 197  
 Db 471995 TGCACCAAGCCCAAGTCCAGGCCCTCAGCGCAGGCTGGGTGTCCAGGCTCGGCGCTCGGC 472054  
 Qy 198 ACAGGAGGCTCAAGTGTACCTGTAGTGGGGGCGCGTGAAGCCCTGAGTTCCTTGGGG 257  
 Db 472055 TCAAGAGCGGCGAAGAACCCCGTGCACCGGATCCCGCGGAGCAGCAGCCGCGGCGCA 472114

Qy 258 AGGCTGTGTCTGAAGAC 274  
 Db 472115 CGGGCGTCTCGACGTC 472131

## RESULT 9

US-10-342-887-1266  
 ; Sequence 1266, Application US/10342887  
 ; Publication No. US20040058340A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dai, Hongyue  
 ; APPLICANT: He, Yudong  
 ; APPLICANT: Linsley, Peter S.  
 ; APPLICANT: Mao, Mao  
 ; APPLICANT: Roberts, Christopher J.  
 ; APPLICANT: Van 't Veer, Laura Johanna  
 ; APPLICANT: Van de Vijver, Marc J.

APPLICANT: Bernards, Rene  
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
; FILE REFERENCE: 9301-188-999  
; CURRENT APPLICATION NUMBER: US/10/342,887  
; CURRENT FILING DATE: 2003-01-15  
; PRIOR APPLICATION NUMBER: 60/298,918  
; PRIOR FILING DATE: 2001-06-18  
; PRIOR APPLICATION NUMBER: 60/380,710  
; PRIOR FILING DATE: 2002-05-14  
; PRIOR APPLICATION NUMBER: 10/172,118  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 2699  
; SEQ ID NO 1266  
; LENGTH: 4180  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-342-887-1266

Query Match 11.1%; Score 42.6; DB 13; Length 4180;  
Best Local Similarity 48.5%; Pred. No. 0.018;  
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
QY 144 AGCCACCAAGTCCAGCCCTCAGCGAGGCTGGGTGTCACAGCCTACTCGACAGGA 203  
Db |||||  
313 AGCCGAGGAGCGCGGTTCAGGACAGGGGCTGGATGGCCGCCACACAGTGGAGG 372  
QY 204 GGCTCAAGCTGTACCTGAGTGGGGCGCGCTGAGGCGCCCTGAGTTCCCTGGGGAGGCTG 263  
Db |||||  
373 AGCTGAAGCGCGCTGAGCGCGGCGAGCTGAGGCGCGCGCTGCTGCGCGTGG 432  
QY 264 TGTCTGAAGACGAATACAGCAAGCTGAGCTGTGATCCAGGAGCTGCTGAGTTCTG 323  
Db |||||  
433 AGCGGAGCTGGCG 492  
QY 324 AGCAGGCTTCTGAGGAGCTGAGTTCTGTCAGAGCCACACCTGAGCACCCTGGAGC 383  
Db |||||  
493 GCCAGAGCAAGGTGAGGCGCTGTACGAGCTGCTGCGGACCGAGGTGCTGGCGCTGCTGC 552  
QY 384 G 384  
Db 553 G 553

RESULT 10  
US-10-172-118-1266  
; Sequence 1266, Application US/10172118  
; Publication No. US20030224374A1  
; GENERAL INFORMATION:  
; APPLICANT: Dai, Hongyue  
; APPLICANT: He, Yudong  
; APPLICANT: Linsley, Peter  
; APPLICANT: Mao, Mao  
; APPLICANT: Roberts, Chris  
; APPLICANT: Van 't Veer, Laura  
; APPLICANT: Van de Vijver, Marc  
; APPLICANT: Bernards, Rene  
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
; FILE REFERENCE: 9301-175-999  
; CURRENT APPLICATION NUMBER: US/10/172,118  
; CURRENT FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: 60/380,770  
; PRIOR FILING DATE: 2002-05-14  
; NUMBER OF SEQ ID NOS: 2699  
; SEQ ID NO 1266  
; LENGTH: 4180  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: NM 006291  
; DATABASE ENTRY DATE: 2001-06-18  
US-10-172-118-1266

Query Match 11.1%; Score 42.6; DB 13; Length 4180;

Best Local Similarity 48.5%; Pred. No. 0.018;  
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
QY 144 AGCCACCAAGTCCAGCCCTCAGCGAGGCTGGGTGTCACAGCCTACTCGACAGGA 203  
Db |||||  
313 AGCCGAGGAGCGCGGTTCAGGACAGGGGCTGGATGGCCGCCACACAGTGGAGG 372  
QY 204 GGCTCAAGCTGTACCTGAGTGGGGCGCGCTGAGGCGCCCTGAGTTCCCTGGGGAGGCTG 263  
Db |||||  
373 AGCTGAAGCGCGCTGAGCGCGGCGAGCTGAGGCGCGCGCTGCTGCGCGTGG 432  
QY 264 TGTCTGAAGACGAATACAGCAAGCTGAGCTGTGATCCAGGAGCTGCTGAGTTCTG 323  
Db |||||  
433 AGCGGAGCTGGCG 492  
QY 324 AGCAGGCTTCTGAGGAGCTGAGTTCTGTCAGAGCCACACCTGAGCACCCTGGAGC 383  
Db |||||  
493 GCCAGAGCAAGGTGAGGCGCTGTACGAGCTGCTGCGGACCGAGGTGCTGGCGCTGCTGC 552  
QY 384 G 384  
Db 553 G 553

RESULT 11  
US-10-439-388-10  
; Sequence 10, Application US/10439388  
; Publication No. US20030228617A1  
; GENERAL INFORMATION:  
; APPLICANT: Aune, Thomas M  
; APPLICANT: Olsen, Nancy J  
; TITLE OF INVENTION: Method for Predicting Autoimmune Disease  
; FILE REFERENCE: 1242/68  
; CURRENT APPLICATION NUMBER: US/10/439,388  
; CURRENT FILING DATE: 2003-05-16  
; PRIOR APPLICATION NUMBER: US 60/381,055  
; PRIOR FILING DATE: 2002-05-16  
; NUMBER OF SEQ ID NOS: 70  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 10  
; LENGTH: 4180  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-439-388-10

Query Match 11.1%; Score 42.6; DB 16; Length 4180;  
Best Local Similarity 48.5%; Pred. No. 0.018;  
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
QY 144 AGCCACCAAGTCCAGCCCTCAGCGAGGCTGGGTGTCACAGCCTACTCGACAGGA 203  
Db |||||  
313 AGCCGAGGAGCGCGGTTCAGGACAGGGGCTGGATGGCCGCCACACAGTGGAGG 372  
QY 204 GGCTCAAGCTGTACCTGAGTGGGGCGCGCTGAGGCGCCCTGAGTTCCCTGGGGAGGCTG 263  
Db |||||  
373 AGCTGAAGCGCGCTGAGCGCGGCGAGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCG 432  
QY 264 TGTCTGAAGACGAATACAGCAAGCTGAGCTGTGATCCAGGAGCTGCTGAGTTCTG 323  
Db |||||  
433 AGCGGAGCTGGCG 492  
QY 324 AGCAGGCTTCTGAGGAGCTGAGTTCTGTCAGAGCCACACCTGAGCACCCTGGAGC 383  
Db |||||  
493 GCCAGAGCAAGGTGAGGCGCTGTACGAGCTGCTGCGGACCGAGGTGCTGGCGCTGCTGC 552  
QY 384 G 384  
Db 553 G 553

RESULT 12  
US-10-776-827-93  
; Sequence 93, Application US/10776827

Publication No. US20040132086A1  
GENERAL INFORMATION:  
APPLICANT: Horwitz, Kathryn  
APPLICANT: Richer, Jennifer  
TITLE OF INVENTION: Progesterone Receptor-Regulated Gene Expression and Methods Related  
TITLE OF INVENTION: Theteto  
FILE REFERENCE: 2848-33  
CURRENT APPLICATION NUMBER: US/10/776,827  
CURRENT FILING DATE: 2004-02-10  
PRIOR APPLICATION NUMBER: US/09/814,915  
PRIOR FILING DATE: 2002-03-21  
PRIOR APPLICATION NUMBER: 60/214,870  
PRIOR FILING DATE: 2000-06-28  
NUMBER OF SEQ ID NOS: 108  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 93  
LENGTH: 4180  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-776-827-93

Query Match 11.1%; Score 42.6; DB 17; Length 4180;  
Best Local Similarity 48.5%; Pred. No. 0.018;  
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
QY 144 AGCCACCAAGTCCAGCCCTCAGCGAGGCTGGGTGTACACAGCCTACTGGACAGGA 203  
DB 313 AGCCCGAGGACGACCGGTTCAGGACAGGCGTGGATGCCCGCCACACAGTGGAGG 372  
QY 204 GGCTCAAGCTGTCACTGAGTGGGGGCGCGCTGAGGCGCCCTGAGTTCCCTGGGGAGGCTG 263  
DB 373 AGCTGAAGCGCGCTGGAGCGCGGCGAGCTGGAGGCGCGCGCTGCTGGCGCTG 432  
QY 264 TGTCTGAAGCGATACAGCAGGCTGAGCTGTGTATCCAGGAGCTCTGAGTTCTG 323  
DB 433 AGCGGAGCTGGCGGCGCGCGCGCGCGCTGTGTATCCAGGAGGAGCTGGTGGCG 492  
QY 324 AGCAGGCTTGTGGAGGAGTGTGAGTTCTGTGACAGACCCACCTGACACCTGAGC 383  
DB 493 GCCAGAGCAGGTGGAGGCGCTGTACGAGCTGTGTGCGGACAGGTGCTGGCGTCTG 552  
QY 384 G 384  
DB 553 G 553

RESULT 13  
US-09-778-927A-16  
Sequence 16, Application US/09778927A  
Patent No. US2002068342A1  
GENERAL INFORMATION:  
APPLICANT: KHOSRAVI, Rami et al.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID AND AMINO ACID SEQUENCES AND NOVEL  
TITLE OF INVENTION: VARIANTS OF ALTERNATIVE SPLICING  
FILE REFERENCE: 2786-0160P  
CURRENT APPLICATION NUMBER: US/09/778,927A  
CURRENT FILING DATE: 2001-02-08  
PRIOR APPLICATION NUMBER: IL 134453  
PRIOR FILING DATE: 2000-02-09  
PRIOR APPLICATION NUMBER: IL135341  
PRIOR FILING DATE: 2000-03-29  
NUMBER OF SEQ ID NOS: 81  
SOFTWARE: Patent in ver. 2.1  
SEQ ID NO 16  
LENGTH: 4282  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(4282)  
OTHER INFORMATION: n = a, c, g, t any unknown or other  
US-09-778-927A-16

Query Match 11.1%; Score 42.6; DB 9; Length 4282;  
Best Local Similarity 48.5%; Pred. No. 0.018;  
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
QY 144 AGCCACCAAGTCCAGCCCTCAGCGAGGCTGGGTGTACACAGCCTACTGGACAGGA 203  
DB 526 AGCCCGAGGACGACCGCGGTTCAGGACAGGCGTGGATGCCCGCCACACAGTGGAGG 585  
QY 204 GGCTCAAGCTGTCACTGAGTGGGGGCGCGCTGAGGCGCCCTGAGTTCCCTGGGGAGGCTG 263  
DB 586 AGCTGAAGCGCGCTGGAGCGCGGCGAGCTGGAGCGCGCGCGCTGCTGGCGCTGG 645  
QY 264 TGTCTGAAGCGATACAGCAGGCTGAGCTGTGTATCCAGGAGCTCTGAGTTCTG 323  
DB 646 AGCGGAGCTGGCGGCGCGCGCGCGCGCTGTGTATCCAGGAGGAGCTGGTGGCGG 705  
QY 324 AGCAGGCTTGTGGAGGAGTGTGAGTTCTGTGACAGCTGTGTGCGGACCAAGTGTGGCGTGTGCTGC 383  
DB 706 GCCAGAGCAGGTGGAGGCGCTGTACGAGCTGTGTGCGGACCAAGTGTGGCGTGTGCTGC 765  
QY 384 G 384  
DB 766 G 766

RESULT 14  
US-10-044-090-668  
Sequence 668, Application US/10044090  
Publication No. US20020137081A1  
GENERAL INFORMATION:  
APPLICANT: Olga Bandman  
TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION  
FILE REFERENCE: PA-0028 US  
CURRENT APPLICATION NUMBER: US/10/044,090  
CURRENT FILING DATE: 2002-01-09  
NUMBER OF SEQ ID NOS: 850  
SOFTWARE: PERL Program  
SEQ ID NO 668  
LENGTH: 4366  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20020137081A1 319136.21  
LOCATION: 3345-3660  
OTHER INFORMATION: a, t, c, g, or other  
US-10-044-090-668

Query Match 11.1%; Score 42.6; DB 14; Length 4366;  
Best Local Similarity 48.5%; Pred. No. 0.018;  
Matches 117; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
QY 144 AGCCACCAAGTCCAGCCCTCAGCGAGGCTGGGTGTACACAGCCTACTGGACAGGA 203  
DB 412 AGCCCGAGGACGACCGCGGTTCAGGACAGGCGTGGATGCCCGCCACACAGTGGAGG 471  
QY 204 GGCTCAAGCTGTCACTGAGTGGGGGCGCGCTGAGGCGCCCTGAGTTCCCTGGGGAGGCTG 263  
DB 472 AGCTGAAGCGCGCTGGAGCGCGGCGAGCTGGAGGCGCGCGCGCTGCTGGCGCTGG 531  
QY 264 TGTCTGAAGCGATACAGCAGGCTGAGCTGTGTATCCAGGAGCTCTGAGTTCTG 323  
DB 532 AGCGGAGCTGGCGGCGCGCGCGCGCGCTGTGTATCCAGGAGGAGCTGGTGGCGG 591  
QY 324 AGCAGGCTTGTGGAGGAGTGTGAGTTCTGTGACAGACCCACCTGACACCTGAGC 383  
DB 592 GCCAGAGCAGGTGGAGGCGCTGTACGAGCTGTGTGCGGACCAAGTGTGGCGTGTGCTGC 651  
QY 384 G 384  
DB 652 G 652

RESULT 15  
US-09-965-528-46  
; Sequence 46, Application US/09965528  
; Publication No. US20020187523A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: TANG, Y. Tom  
; APPLICANT: YUE, Henry  
; APPLICANT: LAL, Preeti  
; APPLICANT: BURFORD, Neil  
; APPLICANT: BANDMAN, Olga  
; APPLICANT: BAUGHN, Mariah R.  
; APPLICANT: AZIMZAI, Valda  
; APPLICANT: LU, Dvung Aina M.  
; APPLICANT: PATTERSON, Chandra  
; TITLE OF INVENTION: EXTRACELLULAR SIGNALING MOLECULES  
; FILE REFERENCE: PF-0701 USA  
; CURRENT APPLICATION NUMBER: US/09/965,528  
; CURRENT FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: 60/134,949  
; PRIOR FILING DATE: 1999-05-19  
; PRIOR APPLICATION NUMBER: 60/144,270  
; PRIOR FILING DATE: 1999-07-15  
; PRIOR APPLICATION NUMBER: 60/146,700  
; PRIOR FILING DATE: 1999-07-30  
; PRIOR APPLICATION NUMBER: 60/157,508  
; PRIOR FILING DATE: 1999-10-04  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: PEARL Program  
; SEQ ID NO 46  
; LENGTH: 1803  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20020187523A1 843193CB1  
US-09-965-528-46

Query Match 10.8%; Score 41.4; DB 9; Length 1803;  
Best Local Similarity 47.5%; Pred. No. 0.037;  
Matches 123; Conservative 0; Mismatches 136; Indels 0; Gaps 0;  
Qy 115 CCACCCACTGGCTGGCTTTCGCGACCAAGCCCAAGTCCAGCCCTCACGGCAGGG 174  
Db 1501 CCAAGGACTGGCTGGGCTGGGGAGGGGAGCCAGATCCCCGAGGGAGGACCCCTGAGGG 1560  
Qy 175 CTGGGTGTACCCAGCCTACCTGGACAGGAGGCTCAAGCTGTCACTGAGTGGGGGGCGC 234  
Db 1561 CCGCGAAGATCCGAGCCGCCAGCTGGGAAGGGGAGGGCGGGTCCCCAGGGGGCGGTGGC 1620  
Qy 235 TGAGGCCCTGAGTTCCTTGGGGAGGCTGTGTCTTAAGCAAGCAATACAAGCAAGGCTGAG 294  
Db 1621 ACAGTGGCCCTTCCCGACGGGTGGCAGGGCCCTGGAGAGGAAGTGTGACCCCTGAT 1680  
Qy 295 CTCTGTGATCCAGGAGCTGCTGAGTTCGACAGGCCCTTCGTGGAGGAGCTGCAGTTCCT 354  
Db 1681 CTCAGGCCACCAAGCTCTGCGGGCTCCAGCGGGGCTCTGAGAGCCCGCTGAAGTCA 1740  
Qy 355 GCAGAGCCACCACTGCAG 373  
Db 1741 GCGACTTAAGGCTTGCAG 1759

Search completed: September 20, 2004, 02:52:12  
Job time : 214.275 secs

**This Page Blank (uspto)**



16	4533	11.0	871	14	US-10-307-019-7	Sequence 7, Appli
17	4529	11.0	871	16	US-10-311-034-20	Sequence 20, Appli
18	3580.5	8.7	26926	9	US-09-759-508B-2	Sequence 2, Appli
19	2629	6.4	4162	15	US-10-023-634-92	Sequence 92, Appli
20	2489	6.0	6642	15	US-10-369-493-5013	Sequence 5013, Ap
21	2447.5	5.9	548	14	US-10-307-019-8	Sequence 8, Appli
22	2415.5	5.9	548	14	US-10-307-019-9	Sequence 9, Appli
23	2324	5.6	3208	15	US-10-210-130-38	Sequence 38, Appli
24	2319	5.6	3268	15	US-10-379-381-2	Sequence 2, Appli
25	2270	5.5	3252	15	US-10-210-130-36	Sequence 36, Appli
26	2203	5.3	3262	15	US-10-379-381-4	Sequence 4, Appli
27	2135	5.2	2380	12	US-10-333-314-18	Sequence 18, Appli
28	2102	5.1	5635	16	US-10-451-168-78	Sequence 78, Appli
29	2085.5	5.1	3186	15	US-10-210-130-34	Sequence 34, Appli
30	2085	5.1	416	15	US-10-108-460A-4389	Sequence 4389, Ap
31	2072.5	5.0	5636	14	US-10-032-189-128	Sequence 128, App
32	2072.5	5.0	5636	15	US-10-120-801-72	Sequence 72, Appli
33	2072.5	5.0	5636	15	US-10-023-634-93	Sequence 93, Appli
34	2072.5	5.0	5636	16	US-10-408-765A-1895	Sequence 1895, Ap
35	2067	5.0	2231	15	US-10-379-381-5	Sequence 5, Appli
36	2067	5.0	2242	16	US-10-408-765A-793	Sequence 793, App
37	2048	5.0	5198	12	US-10-346-863-34	Sequence 34, Appli
38	2048	5.0	5198	15	US-10-120-801-75	Sequence 75, Appli
39	2048	5.0	5198	15	US-10-369-493-6858	Sequence 6858, Ap
40	2048	5.0	5198	15	US-10-369-493-6860	Sequence 6860, Ap
41	2045	5.0	390	12	US-10-425-114-37530	Sequence 37530, A
42	2043.5	5.0	5175	15	US-10-120-801-74	Sequence 74, Appli
43	2043.5	5.0	5175	15	US-10-369-493-6859	Sequence 6859, Ap
44	2043.5	5.0	5175	15	US-10-369-493-6861	Sequence 6861, Ap
45	1840	4.5	3931	15	US-10-120-801-18	Sequence 18, Appli

ALIGNMENTS

RESULT 1

US-10-077-130-5  
; Sequence 5, Application US/10077130  
; Publication NO. US20020168742A1  
; GENERAL INFORMATION:  
; APPLICANT: Kapeller-Libermann, Rosana  
; APPLICANT: Acton, Susan L.  
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family  
; TITLE OF INVENTION: Members and Uses Thereof  
; FILE REFERENCE: MPI2001-047P1RCP1(M)  
; CURRENT APPLICATION NUMBER: US/10/077,130  
; CURRENT FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: 60/269201  
; PRIOR FILING DATE: 2001-02-15  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 7968  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-077-130-5

Query Match 100.0%; Score 41273; DB 13; Length 7968;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 7968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MDQFSGAPFLTEPKAFVSVGKDATLSQIVGNPTPOVSWKDDQPVTAGARFLAQ	60
DB	1	MDQFSGAPFLTEPKAFVSVGKDATLSQIVGNPTPOVSWKDDQPVTAGARFLAQ	60
QY	61	DGDLRLRLDLDALGDSQYVCRARNATGEAFVAVGLQVDAEAAACAEQPHLLRPTSIR	120
DB	61	DGDLRLRLDLDALGDSQYVCRARNATGEAFVAVGLQVDAEAAACAEQPHLLRPTSIR	120
QY	121	VREGSEATFCRVGSGPRPAVSWKDGRRIGCEPDGPRVVEELGEASALRTAARPDGG	180
DB	121	VREGSEATFCRVGSGPRPAVSWKDGRRIGCEPDGPRVVEELGEASALRTAARPDGG	180

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 11:36:00 ; Search time 480 Seconds  
(without alignments)  
5323.448 Million cell updates/sec

Title: US-10-077-130-5  
Perfect score: 41273  
Sequence: 1 MDQFSGAPFLTEPKAFV.....RNKKRRALLYKHNLAQVR 7968

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41273	100.0	7968	13	US-10-077-130-5
2	32134	77.9	6620	15	US-10-080-334-290
3	32134	77.9	6620	16	US-10-408-765A-2291
4	13710	33.2	2630	13	US-10-077-130-2
5	13528	32.8	2596	14	US-10-307-019-6
6	10519.5	25.5	4691	15	US-10-093-463-72
7	10506	25.5	4675	15	US-10-093-463-74
8	9834.5	23.8	2328	16	US-10-476-397-4
9	8423	20.4	1665	9	US-09-858-664A-2
10	8423	20.4	1665	12	US-10-697-263-2
11	8410	20.4	1665	12	US-10-415-011-22
12	8407	20.4	1618	12	US-10-182-243-46
13	8403	20.4	1610	14	US-10-307-019-4
14	7073	17.1	1351	14	US-10-307-019-1
15	6619.5	16.0	1596	16	US-10-408-765A-992

QY 181 TYEVRAENPLGAASAAAALVVDSDAADTASRPGTSTAALLAHLORREARAEAPASPP 240  
DB 181 TYEVRAENPLGAASAAAALVVDSDAADTASRPGTSTAALLAHLORREARAEAPASPP 240  
QY 241 STGTRTCTVTGEGHARLSCVVTGEPKPEVTWKXDGQLVTEGRRHVYVEDAQENFVLKILF 300  
DB 241 STGTRTCTVTGEGHARLSCVVTGEPKPEVTWKXDGQLVTEGRRHVYVEDAQENFVLKILF 300  
QY 301 CKOSDRGLYCTASNLVGQYSSVLVVRBPAPVFFKRGLODLVREKESATFLCEVPQPS 360  
DB 301 CKOSDRGLYCTASNLVGQYSSVLVVRBPAPVFFKRGLODLVREKESATFLCEVPQPS 360  
QY 361 TEAAWFKBEETRLWASAKYGEESGTERRLTVRNVSDDDAVYICETPEGSRTVAELAVOG 420  
DB 361 TEAAWFKBEETRLWASAKYGEESGTERRLTVRNVSDDDAVYICETPEGSRTVAELAVOG 420  
QY 421 NLLRKLPRKTAVRVGDTPAMFCVELAVPVPVHMLRNQEEVVAGRVVAISAEGRTHLTIS 480  
DB 421 NLLRKLPRKTAVRVGDTPAMFCVELAVPVPVHMLRNQEEVVAGRVVAISAEGRTHLTIS 480  
QY 481 OCCLEDVGOAVFAGDCOTSTRFCVSAPRPPLOPPVDVVKARMESSVILSNPPHGE 540  
DB 481 OCCLEDVGOAVFAGDCOTSTRFCVSAPRPPLOPPVDVVKARMESSVILSNPPHGE 540  
QY 541 RPTVIDGYLVKXKLGITYTWIRCHEAEWATPELTIVADVAEENGFQFRVSALNSFGQSPY 600  
DB 541 RPTVIDGYLVKXKLGITYTWIRCHEAEWATPELTIVADVAEENGFQFRVSALNSFGQSPY 600  
QY 601 LEPPGTVHLAPKLVNTPKAVQAVEGVTFSVDLTVASAGWFLDGOALKASSVYIEH 660  
DB 601 LEPPGTVHLAPKLVNTPKAVQAVEGVTFSVDLTVASAGWFLDGOALKASSVYIEH 660  
QY 661 CDRTRHTLTIREVPASLHGQALFXFANGIESSIRMEVRAAPGLTANKPPAAAAREVLARL 720  
DB 661 CDRTRHTLTIREVPASLHGQALFXFANGIESSIRMEVRAAPGLTANKPPAAAAREVLARL 720  
QY 721 HEBQALLAEISDQAAAATWIKOORTLSPGPKYEVQASAGRRVLLVDVARDAGLYECVS 780  
DB 721 HEBQALLAEISDQAAAATWIKOORTLSPGPKYEVQASAGRRVLLVDVARDAGLYECVS 780  
QY 781 RGGRIAYQLSVQGLARFLHKDMAGSCVDAVAGGPAQFECETSEAHVHVHWYKDGMBELGHS 840  
DB 781 RGGRIAYQLSVQGLARFLHKDMAGSCVDAVAGGPAQFECETSEAHVHVHWYKDGMBELGHS 840  
QY 841 GERFLQEDVGTGRHLVAATVTRDEGTYSCRVGEDSVDFRLRVSEPKVVFABKEQLARRKL 900  
DB 841 GERFLQEDVGTGRHLVAATVTRDEGTYSCRVGEDSVDFRLRVSEPKVVFABKEQLARRKL 900  
QY 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVCMGATGCTRLVLVQOAGQADAGY 960  
DB 901 QAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVCMGATGCTRLVLVQOAGQADAGY 960  
QY 961 SCAGGQRLSFHLDVKEPKVVFADQVVAHSEVOAEAGANATLSCEVAQAQAEVWYKDGK 1020  
DB 961 SCAGGQRLSFHLDVKEPKVVFADQVVAHSEVOAEAGANATLSCEVAQAQAEVWYKDGK 1020  
QY 1021 KLSSSLVHVYBAKCRRLVLVQOAGTKTDAGDYSCAARGQVSRFLHTBPKXWFAKEQSV 1080  
DB 1021 KLSSSLVHVYBAKCRRLVLVQOAGTKTDAGDYSCAARGQVSRFLHTBPKXWFAKEQSV 1080  
QY 1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSSSSKVGMVKGCTRLVLVQAGKAD 1140  
DB 1081 HNEVQAEAGASAMLSCEVAQAQTEVTWYKDGKLSSSSKVGMVKGCTRLVLVQAGKAD 1140  
QY 1141 AGEYSCBAGQORVSFHLHITTEPKGVFAKEQSVNEVOAEBAGTTAMLSCVAQOQTEVTWY 1200  
DB 1141 AGEYSCBAGQORVSFHLHITTEPKGVFAKEQSVNEVOAEBAGTTAMLSCVAQOQTEVTWY 1200  
QY 1201 KDGKLSSSSKVMEVKGCTRLVLVQOAGKADAGEYSCBAGQORVSFQHLHITTEPKAVFAK 1260  
DB 1201 KDGKLSSSSKVMEVKGCTRLVLVQOAGKADAGEYSCBAGQORVSFQHLHITTEPKAVFAK 1260  
QY 1261 EQLVHNEVRTAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVRIEACGCKWRQLVQQA 1320

DB 1261 EQLVHNEVRTAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVRIEACGCKWRQLVQQA 1320  
QY 1321 GOADAGEYTCBAGQORLSFHLVDSEPKAVFAKSOAHKRVQAEAGATATSCVEQAQOTE 1380  
DB 1321 GOADAGEYTCBAGQORLSFHLVDSEPKAVFAKSOAHKRVQAEAGATATLSCEVAQOTE 1380  
QY 1381 VTWYKDGKLSSSSKVMEAVGCTRLVLVQOACQADTGEYSCBAGGQRLSFSLDVABPKV 1440  
DB 1381 VTWYKDGKLSSSSKVMEAVGCTRLVLVQOACQADTGEYSCBAGGQRLSFSLDVABPKV 1440  
QY 1441 VFAKEQPVHREVOAQAGASTTSCVEQAQOTEVWYKDGKLSFSSKVRMEAVGCTRLV 1500  
DB 1441 VFAKEQPVHREVOAQAGASTTSCVEQAQOTEVWYKDGKLSFSSKVRMEAVGCTRLV 1500  
QY 1501 VQOAGQADAGEYSCBAGQORLSFHLHVAEPKAVFAKEQAPASREVOBAGTSATLSCEVAQ 1560  
DB 1501 VQOAGQADAGEYSCBAGQORLSFHLHVAEPKAVFAKEQAPASREVOBAGTSATLSCEVAQ 1560  
QY 1561 AQTEVTWYKDGKLSSSSKVMEAVGCTRLVLVQOAGQADAGEYSCBAGQORLSFHLHVA 1620  
DB 1561 AQTEVTWYKDGKLSSSSKVMEAVGCTRLVLVQOAGQADAGEYSCBAGQORLSFHLHVA 1620  
QY 1621 EPKVVFAKEQPAHREVOAEBAGASATLSCEVAQAQOTEVTWYKDGKLSSSSKVRVEAVGCT 1680  
DB 1621 EPKVVFAKEQPAHREVOAEBAGASATLSCEVAQAQOTEVTWYKDGKLSSSSKVRVEAVGCT 1680  
QY 1681 RRLVVQOAGQADAGEYSCBAGGQRLSFRLHVALEPOISERPCCRRERPLVVKHEHDIILTA 1740  
DB 1681 RRLVVQOAGQADAGEYSCBAGGQRLSFRLHVALEPOISERPCCRRERPLVVKHEHDIILTA 1740  
QY 1741 TLATPSAATVTLWKDGVIEIRRSKHETASQGDTHTLTVHGAQVLDLSAIYSCRVAEGQDF 1800  
DB 1741 TLATPSAATVTLWKDGVIEIRRSKHETASQGDTHTLTVHGAQVLDLSAIYSCRVAEGQDF 1800  
QY 1801 PVQVEEVAAKFCRLLEPVCGLGTTVTLACELSPACAEVWVRGNTOPRVKRFQWVAEG 1860  
DB 1801 PVQVEEVAAKFCRLLEPVCGLGTTVTLACELSPACAEVWVRGNTOPRVKRFQWVAEG 1860  
QY 1861 PVRSLTVLGRAEDAGEYVCESEDDHTSAQLTVSVPVVKFMSGLSTVVAEAGEATFQC 1920  
DB 1861 PVRSLTVLGRAEDAGEYVCESEDDHTSAQLTVSVPVVKFMSGLSTVVAEAGEATFQC 1920  
QY 1921 VVSPSDVAVVWFRGALLQPSKFAISQSGASHLSTISDLVLEDAQOITVVEAGASSAA 1980  
DB 1921 VVSPSDVAVVWFRGALLQPSKFAISQSGASHLSTISDLVLEDAQOITVVEAGASSAA 1980  
QY 1981 LRVEAPVLFPKKLEPQTVBEERSVTLLEBELTRPPELRTWRNATALAPGNVEIHAEGA 2040  
DB 1981 LRVEAPVLFPKKLEPQTVBEERSVTLLEBELTRPPELRTWRNATALAPGNVEIHAEGA 2040  
QY 2041 RHRLVLNVGFADRGFGCETPDQKTOAKLTVMRQVRLYRGICAVEARCGGTATVEQL 2100  
DB 2041 RHRLVLNVGFADRGFGCETPDQKTOAKLTVMRQVRLYRGICAVEARCGGTATVEQL 2100  
QY 2101 SHADVGSWTRDGLRFQCGPTCHLAVRGPMTLTLGLRPEDSGLMVFKAEGVHTSARLV 2160  
DB 2101 SHADVGSWTRDGLRFQCGPTCHLAVRGPMTLTLGLRPEDSGLMVFKAEGVHTSARLV 2160  
QY 2161 VTELVPVSFRPLQDVVTTEKVTLECELSRPVNDVRLWKDGVBLRAGKTMAIAAQACR 2220  
DB 2161 VTELVPVSFRPLQDVVTTEKVTLECELSRPVNDVRLWKDGVBLRAGKTMAIAAQACR 2220  
QY 2221 SLTIYRCFEFADQGVYCDADHAQSSASVKVQGRYTLIYRRVLAEDAGEIQFVAENAESR 2280  
DB 2221 SLTIYRCFEFADQGVYCDADHAQSSASVKVQGRYTLIYRRVLAEDAGEIQFVAENAESR 2280  
QY 2281 AQLRVKELPVLTVRPLRDKTIAMEKHGVLBCQVSRASQVRFKSGOELQFGPKYELVSD 2340  
DB 2281 AQLRVKELPVLTVRPLRDKTIAMEKHGVLBCQVSRASQVRFKSGOELQFGPKYELVSD 2340  
QY 2341 GLYRKLIIISDVHAEDEDTYTCADGDKNTSAQFFVEBOSITIVGLQDQVTVMEPAPAFEC 2400

Db	2341	GLYRKLIISDVHAEDEYTCADGDKTSAQFVEEQSIITVRGLQDVTVMEPAPAFEC	2400
Qy	2401	ETSIPIVVRPPKWLIGKTVLQAGNVGLQEGTQVHRLMLRTCTMTGPHFTVVKSRSSA	2460
Db	2401	ETSIPIVVRPPKWLIGKTVLQAGNVGLQEGTQVHRLMLRTCTMTGPHFTVVKSRSSA	2460
Qy	2461	RLVVSDDIPVLTPLPKTGRELOSVLSCDFRPAKAVQWYKDDTPLSPSEKFKMSLEG	2520
Db	2461	RLVVSDDIPVLTPLPKTGRELOSVLSCDFRPAKAVQWYKDDTPLSPSEKFKMSLEG	2520
Qy	2521	QMAELRIILRLPADAGVYRCQAGSAHSSTEVTVREAREVTVTGPIQDAEATEEGWASFSCE	2580
Db	2521	QMAELRIILRLPADAGVYRCQAGSAHSSTEVTVREAREVTVTGPIQDAEATEEGWASFSCE	2580
Qy	2581	LSHEDDEEWELNGMPLYNDSFHEISHKGRHTLVLSKIQPADAGIVRASLKVSTSAKL	2640
Db	2581	LSHEDDEEWELNGMPLYNDSFHEISHKGRHTLVLSKIQPADAGIVRASLKVSTSAKL	2640
Qy	2641	EVVRKPVVFLKALDDLSAEERGTALQCEVSDPEAHVVRKDGQVQLGSPDKYDLFHTAGT	2700
Db	2641	EVVRKPVVFLKALDDLSAEERGTALQCEVSDPEAHVVRKDGQVQLGSPDKYDLFHTAGT	2700
Qy	2701	RGLVVHDVSPEDAGLYTCHVGSEETRAVRVHDLHVGITKELKTMEVLEBESCSFECVLS	2760
Db	2701	RGLVVHDVSPEDAGLYTCHVGSEETRAVRVHDLHVGITKELKTMEVLEBESCSFECVLS	2760
Qy	2761	HESASDPAMWTVGGKTVGSSSRFOATROGRXYILVWREAPSADAGEVVFVSRGLTSKASL	2820
Db	2761	HESASDPAMWTVGGKTVGSSSRFOATROGRXYILVWREAPSADAGEVVFVSRGLTSKASL	2820
Qy	2821	IVRERPAIILKPLEDQWVAPGEDVELRCELRSAGTPVHMLKDKAIRKSQKYDVVCEGTM	2880
Db	2821	IVRERPAIILKPLEDQWVAPGEDVELRCELRSAGTPVHMLKDKAIRKSQKYDVVCEGTM	2880
Qy	2881	AMLVIRGASLKDAGEYTCVEASKSTASLHVEEKANCFTBELTNLQVEEKGTAFTCKTE	2940
Db	2881	AMLVIRGASLKDAGEYTCVEASKSTASLHVEEKANCFTBELTNLQVEEKGTAFTCKTE	2940
Qy	2941	HPAATVTVWRKGLLELRASGKHQPSQEGTLTLTTSALEKADSDTYTCDIGQAQRAQLLV	3000
Db	2941	HPAATVTVWRKGLLELRASGKHQPSQEGTLTLTTSALEKADSDTYTCDIGQAQRAQLLV	3000
Qy	3001	QGRVHIITIEDLVVDVQEGSSATFCRISPANYEPVHWFDKTPIHANELNEIDAQPGGY	3060
Db	3001	QGRVHIITIEDLVVDVQEGSSATFCRISPANYEPVHWFDKTPIHANELNEIDAQPGGY	3060
Qy	3061	HVLTIRQLALKDSGTIFYEAGDQASALRVTEKPSVFSRELTDATITEGEDTLTVCETS	3120
Db	3061	HVLTIRQLALKDSGTIFYEAGDQASALRVTEKPSVFSRELTDATITEGEDTLTVCETS	3120
Qy	3121	TCDIPMCWTQDKTLRGARCOLSHEGHAQLLITGATLQDSGRYKCEAGGACSSIVRV	3180
Db	3121	TCDIPMCWTQDKTLRGARCOLSHEGHAQLLITGATLQDSGRYKCEAGGACSSIVRV	3180
Qy	3181	HARPVRFQALKDLEVLGGAAATLRCVLSVAAPVKWCYGNVLRPGDKYSLROEGAMLE	3240
Db	3181	HARPVRFQALKDLEVLGGAAATLRCVLSVAAPVKWCYGNVLRPGDKYSLROEGAMLE	3240
Qy	3241	LVRNLRPDQSGRYSCSGFDQTTSATLTVTALPAQFIGIKLRNKEATEGATATLRCLSKT	3300
Db	3241	LVRNLRPDQSGRYSCSGFDQTTSATLTVTALPAQFIGIKLRNKEATEGATATLRCLSKT	3300
Qy	3301	APVEWRKSGSETLRDGRYCLRDGAMCELOIRGLAMVDAAEYSCVCGEERTSASLTIRPM	3360
Db	3301	APVEWRKSGSETLRDGRYCLRDGAMCELOIRGLAMVDAAEYSCVCGEERTSASLTIRPM	3360
Qy	3361	PAHFIRGRLRHOBESIGATATLRCLSKAAPVEWRKGRESLRDGRHSLRQDGAVCELOIC	3420
Db	3361	PAHFIRGRLRHOBESIGATATLRCLSKAAPVEWRKGRESLRDGRHSLRQDGAVCELOIC	3420
Qy	3421	GLAVADAGEYSCVCGEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLWCLSKVAPVE	3480
Db	3421	GLAVADAGEYSCVCGEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLWCLSKVAPVE	3480
Qy	3481	WRKGPENLRDGRYTLRQEGTRCELOICGLAMADAGEVLCVCGOERTSATLTIRALPARF	3540
Db	3481	WRKGPENLRDGRYTLRQEGTRCELOICGLAMADAGEVLCVCGOERTSATLTIRALPARF	3540
Qy	3541	IEDVKNQEARREGATAVLOCELNSAAPVEWRKSGSETLRDGRYSLRQDGTKCELOIRGLAM	3600
Db	3541	IEDVKNQEARREGATAVLOCELNSAAPVEWRKSGSETLRDGRYSLRQDGTKCELOIRGLAM	3600
Qy	3601	AOTGEYSCVCGOERTSAMLTVRALPIKFTTEGLRNEEAETEGATAVLRCLSKNAPVEWKKG	3660
Db	3601	AOTGEYSCVCGOERTSAMLTVRALPIKFTTEGLRNEEAETEGATAVLRCLSKNAPVEWKKG	3660
Qy	3661	HETLRDGRHSLRQDGARCELOIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPSFIEGL	3720
Db	3661	HETLRDGRHSLRQDGARCELOIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPSFIEGL	3720
Qy	3721	RNEEAETEGATATLWCLSKAAPVEWRKSGSETLRDGRHSLRQDGRCELOIRGLAVWDAG	3780
Db	3721	RNEEAETEGATATLWCLSKAAPVEWRKSGSETLRDGRHSLRQDGRCELOIRGLAVWDAG	3780
Qy	3781	EYSCVCGOERTSATLTVRALPARFIEDVKNQEARREGATAVLOCELNSKAAPVEWRKSGSETL	3840
Db	3781	EYSCVCGOERTSATLTVRALPARFIEDVKNQEARREGATAVLOCELNSKAAPVEWRKSGSETL	3840
Qy	3841	RGDRYSLRQDGTTRCELOIRGLSVADTGEYSCVCGOERTSATLTVRAPQVFPREPQLQSLQ	3900
Db	3841	RGDRYSLRQDGTTRCELOIRGLSVADTGEYSCVCGOERTSATLTVRAPQVFPREPQLQSLQ	3900
Qy	3901	AREGSTATLQCELSSEPTATVWMSKGLQLOQANGREPRELOGCTAEILVLODREDTGEVT	3960
Db	3901	AREGSTATLQCELSSEPTATVWMSKGLQLOQANGREPRELOGCTAEILVLODREDTGEVT	3960
Qy	3961	CTCGSQAATSATLTVAAPVRFLELQHQVEDEGGTAHLCCELSRAGASVEMWKGSLQPLF	4020
Db	3961	CTCGSQAATSATLTVAAPVRFLELQHQVEDEGGTAHLCCELSRAGASVEMWKGSLQPLF	4020
Qy	4021	CARYQWQDGAALLLVRGVEQEDAGDYTCDTGHTQSMASLSVRVDRPFKTRLSQLEOE	4080
Db	4021	CARYQWQDGAALLLVRGVEQEDAGDYTCDTGHTQSMASLSVRVDRPFKTRLSQLEOE	4080
Qy	4081	TGDIARLCCQLSDAESGAVVQWLKEGVELHAGPKYEMRSQGATRELLIHQLEAKDTGEYA	4140
Db	4081	TGDIARLCCQLSDAESGAVVQWLKEGVELHAGPKYEMRSQGATRELLIHQLEAKDTGEYA	4140
Qy	4141	CVTGGOKTAASLRVTEPEVTIVRGLVDAEVTADEVFEFSCEVSRAGATGVQWCLQGLPLQ	4200
Db	4141	CVTGGOKTAASLRVTEPEVTIVRGLVDAEVTADEVFEFSCEVSRAGATGVQWCLQGLPLQ	4200
Qy	4201	SNEVTEVAVRDCRIHTLRKGYTPEDAGTVSHLGNHASSAQLTVREAPVTILEPLQDYO	4260
Db	4201	SNEVTEVAVRDCRIHTLRKGYTPEDAGTVSHLGNHASSAQLTVREAPVTILEPLQDYO	4260
Qy	4261	LSEGDQASFCRLSRASGQEARWALGCVPLQANEMNDITVEQGTLLHLLTHKVTLEDACT	4320
Db	4261	LSEGDQASFCRLSRASGQEARWALGCVPLQANEMNDITVEQGTLLHLLTHKVTLEDACT	4320
Qy	4321	VSHVGTCSSEAOQLKVTAKNTVVRGLNVEALEEGEALFECQLSQEVAHAHTWLLDDEPV	4380
Db	4321	VSHVGTCSSEAOQLKVTAKNTVVRGLNVEALEEGEALFECQLSQEVAHAHTWLLDDEPV	4380
Qy	4381	RTSENAEVVFFENGLRHLLKLRPQDSCRVTFLAGDMVTSAFLTIVRGWRLEILEPLKN	4440
Db	4381	RTSENAEVVFFENGLRHLLKLRPQDSCRVTFLAGDMVTSAFLTIVRGWRLEILEPLKN	4440
Qy	4441	AAVRAGAQAARFTCTLSEAVPVGEASWYINGAAVQPDSDSWTVTADGSHOALLLRSAQPHH	4500
Db	4441	AAVRAGAQAARFTCTLSEAVPVGEASWYINGAAVQPDSDSWTVTADGSHOALLLRSAQPHH	4500
Qy	4501	AGEVTFACRDVAVASARLTVLGLPDPPEDEAVVAHSSHTVTLSWAAPMSDGGGLCGYRVE	4560
Db	4501	AGEVTFACRDVAVASARLTVLGLPDPPEDEAVVAHSSHTVTLSWAAPMSDGGGLCGYRVE	4560

QY	4561	VKEGATGOWRLCHELVPGPEVCVGLAPGETYRFRVAAGVGVGAGEPVHLPTQTVRLAEPP	4620	Db	5641	LVRTPTKSSPRQGWSPAYLDRRLKLSPEWGAEEPEFGEAVSEDEYKARLSSVIOE	5700
Db	4561	VKEGATGOWRLCHELVPGPEVCVGLAPGETYRFRVAAGVGVGAGEPVHLPTQTVRLAEPP	4620	QY	5701	LLSSQAFVVEELQFLOSHLQHLERCPHVPITAVAGQKAVIPRNVRDIDGRFHSSFFQELQO	5760
QY	4621	KPVPPQPSAPSRQVAAAGEDVSLLEVVAEAGEVIVHKGMERIOPGRFEFVSGROQML	4680	Db	5701	LLSSQAFVVEELQFLOSHLQHLERCPHVPITAVAGQKAVIPRNVRDIDGRFHSSFFQELQO	5760
Db	4621	KPVPPQPSAPSRQVAAAGEDVSLLEVVAEAGEVIVHKGMERIOPGRFEFVSGROQML	4680	QY	5761	CTDDDDVAMCFIKNOAAFEQYLFLVGRVQAEVVVSTAIOEFYKXAEALLAGDPSQP	5820
QY	4681	VIKGFTAEDQGEYHCGLAAGSCIPAAATFOVALSPASVDEAPQPSLPEAAQEGDLHLW	4740	Db	5761	CTDDDDVAMCFIKNOAAFEQYLFLVGRVQAEVVVSTAIOEFYKXAEALLAGDPSQP	5820
Db	4681	VIKGFTAEDQGEYHCGLAAGSCIPAAATFOVALSPASVDEAPQPSLPEAAQEGDLHLW	4740	QY	5821	PPPLQHYLEQVERVQRYQALLKELIRNKARNRQNCALLQAYAVVSALPQAEKLUHV	5880
QY	4741	EALARKRMSREPTLDSISELPEEDGRSQRUPQAEVAPDLSSEGYSTADELARTGDADL	4800	Db	5821	PPPLQHYLEQVERVQRYQALLKELIRNKARNRQNCALLQAYAVVSALPQAEKLUHV	5880
Db	4741	EALARKRMSREPTLDSISELPEEDGRSQRUPQAEVAPDLSSEGYSTADELARTGDADL	4800	QY	5881	SLMENYPTGLEALGEPPIRQGHFIWEGAPGARMKWHNRHVFIFRNHLVICKPRRDSRT	5940
QY	4801	SHTSSDDESRACTPSLVTLKAGRPCTSPILASKVGAAPSVKPOQOQEPPLAARPLG	4860	Db	5881	SLMENYPTGLEALGEPPIRQGHFIWEGAPGARMKWHNRHVFIFRNHLVICKPRRDSRT	5940
Db	4801	SHTSSDDESRACTPSLVTLKAGRPCTSPILASKVGAAPSVKPOQOQEPPLAARPLG	4860	QY	5941	DTVSVFRNMKLSIDLDNDQVEGDDRAFEVWQREDSVRKYLLOARTAILKSWKKEIC	6000
QY	4861	DLSTKDLGDPMSDKAAVKIOAFKGYKVRKEMKQOEGMFHSHTFGDTEAQVGDALRECV	4920	Db	5941	DTVSVFRNMKLSIDLDNDQVEGDDRAFEVWQREDSVRKYLLOARTAILKSWKKEIC	6000
Db	4861	DLSTKDLGDPMSDKAAVKIOAFKGYKVRKEMKQOEGMFHSHTFGDTEAQVGDALRECV	4920	QY	6001	GIQORLALPVWRPDPFEBELADCTAELGETVKLACRVGTGTPKVISWYKDKGKAVQDPDH	6060
QY	4921	VASRADVRAWLKDGVELTDGRHHHIDQLDGTCSLIAGLDRADAGCYTCQVSNKFGQV	4980	Db	6001	GIQORLALPVWRPDPFEBELADCTAELGETVKLACRVGTGTPKVISWYKDKGKAVQDPDH	6060
Db	4921	VASRADVRAWLKDGVELTDGRHHHIDQLDGTCSLIAGLDRADAGCYTCQVSNKFGQV	4980	QY	6061	ILIEDPDGSCALILDSLTGVDSGQWMCFAAAGNCSTLGLKILVQVPPRFVKNVRAAPFV	6120
QY	4981	THSACVVVSGSESAESSGGELDDAFRRARRLHRLFRTKSPAESVDEBLFSADEGPA	5040	Db	6061	ILIEDPDGSCALILDSLTGVDSGQWMCFAAAGNCSTLGLKILVQVPPRFVKNVRAAPFV	6120
Db	4981	THSACVVVSGSESAESSGGELDDAFRRARRLHRLFRTKSPAESVDEBLFSADEGPA	5040	QY	6121	EGEDAQFTCTTEGAPYPOIRWYKDGALLTGNKFTLSEPRSGLLVLVIRAAKEDLGLY	6180
QY	5041	EPEPADQTVREDEHFCIRFEALTEARQAVTRFQEMFATLGIVKILVEQOPRREM	5100	Db	6121	EGEDAQFTCTTEGAPYPOIRWYKDGALLTGNKFTLSEPRSGLLVLVIRAAKEDLGLY	6180
Db	5041	EPEPADQTVREDEHFCIRFEALTEARQAVTRFQEMFATLGIVKILVEQOPRREM	5100	QY	6181	ECELNRLGASARASAEIRIQSPMLQAOCHREQLVAAVEDTTLERADQEVTSVLKLLG	6240
QY	5101	CISKETPAPVPPPEPLSLTSDAAPVPLTELQNOEQVQDGYPSFDCVVTGQPMPSVRWF	5160	Db	6181	ECELNRLGASARASAEIRIQSPMLQAOCHREQLVAAVEDTTLERADQEVTSVLKLLG	6240
Db	5101	CISKETPAPVPPPEPLSLTSDAAPVPLTELQNOEQVQDGYPSFDCVVTGQPMPSVRWF	5160	QY	6241	PKAPGPGTDLTGCPGCPGAPALQETGSPVGTSEAPVPPRVPOLHHEGPEOPE	6300
QY	5161	KDGLLEDDHYMNEQOQGHQIITAVVPADMVGVYRCLEANSMSGVSTKAEIRVDLTS	5220	Db	6241	PKAPGPGTDLTGCPGCPGAPALQETGSPVGTSEAPVPPRVPOLHHEGPEOPE	6300
Db	5161	KDGLLEDDHYMNEQOQGHQIITAVVPADMVGVYRCLEANSMSGVSTKAEIRVDLTS	5220	QY	6301	AIARAQEVTVPIRMEGAAMPAGAGTCELLVDVSHVVRVETTTQTYTQAIDTHTARPPSMQ	6360
QY	5221	TDYTAADATESSSYFSAQGYLSSREOEGTESITDQGLPQVVEELRDLOVAPGTRLAKF	5280	Db	6301	AIARAQEVTVPIRMEGAAMPAGAGTCELLVDVSHVVRVETTTQTYTQAIDTHTARPPSMQ	6360
Db	5221	TDYTAADATESSSYFSAQGYLSSREOEGTESITDQGLPQVVEELRDLOVAPGTRLAKF	5280	QY	6361	VTIEDVQAOQTGTAQFAEALIEGDPQPSVTWYKDSVOLVDSTRLSQOQEGTYSVLARHA	6420
QY	5281	QLKVGYPAPRLYKFGKOPITASAHIRMTGKILHLEIISVTRDSGQYAYIISNMG	5340	Db	6361	VTIEDVQAOQTGTAQFAEALIEGDPQPSVTWYKDSVOLVDSTRLSQOQEGTYSVLARHA	6420
Db	5281	QLKVGYPAPRLYKFGKOPITASAHIRMTGKILHLEIISVTRDSGQYAYIISNMG	5340	QY	6421	SKDAGVYVCLAQNTGGQVLCKAELLVLGGDNEPDSQSHRKLHSFYEVKEEIGRGVFG	6480
QY	5341	AAYSARLLVRGPEDEPEKPADSVHEQLVPPMLERFTPKKXKGSITFVSXVEGRPVP	5400	Db	6421	SKDAGVYVCLAQNTGGQVLCKAELLVLGGDNEPDSQSHRKLHSFYEVKEEIGRGVFG	6480
Db	5341	AAYSARLLVRGPEDEPEKPADSVHEQLVPPMLERFTPKKXKGSITFVSXVEGRPVP	5400	QY	6481	FVKRVQHGKGNKILCAAKIPIPLSRTRAQAYBERDILAAALSHPLVTGLLDQETKTLILI	6540
QY	5401	TVHMLREABEGVWIGPDPGTGYVASSAQHSLVLDVGRQHQGTTCIASNAAGALC	5460	Db	6481	FVKRVQHGKGNKILCAAKIPIPLSRTRAQAYBERDILAAALSHPLVTGLLDQETKTLILI	6540
Db	5401	TVHMLREABEGVWIGPDPGTGYVASSAQHSLVLDVGRQHQGTTCIASNAAGALC	5460	QY	6541	LELCSSELLDRLYKGVVTEAEVKVYIQQVLEGLVHLSHGVHLHLDIKPSNIMLVHPAR	6600
QY	5461	SASLVSGLPKVEQEKVKEALISTFLOQTQTAISAOGLTAGFADLGGORKEEPLAAKE	5520	Db	6541	LELCSSELLDRLYKGVVTEAEVKVYIQQVLEGLVHLSHGVHLHLDIKPSNIMLVHPAR	6600
Db	5461	SASLVSGLPKVEQEKVKEALISTFLOQTQTAISAOGLTAGFADLGGORKEEPLAAKE	5520	QY	6601	EDIKICDFGPAQNTTAPAELOFSQSGPEFVSPEIIOQNPVSEASDIWAMGVISLSTCS	6660
QY	5521	ALGHLSLAEVTEBFLQKLSQITEMVSAKITQAKLVQPGDSDSDSKTPSAPRGRSR	5580	Db	6601	EDIKICDFGPAQNTTAPAELOFSQSGPEFVSPEIIOQNPVSEASDIWAMGVISLSTCS	6660
Db	5521	ALGHLSLAEVTEBFLQKLSQITEMVSAKITQAKLVQPGDSDSDSKTPSAPRGRSR	5580	QY	6661	SPFAGESDRAILLNVLEGRVSWSPMAHILSEDAKDFIKATLQAPQARSAAOCLSHPW	6720
QY	5581	PSSSIQFSSSSSEBDGARGEIFDIYVVTADYLPILGAFQDAILTREQOYVEVLDAHPLEW	5640	Db	6661	SPFAGESDRAILLNVLEGRVSWSPMAHILSEDAKDFIKATLQAPQARSAAOCLSHPW	6720
Db	5581	PSSSIQFSSSSSEBDGARGEIFDIYVVTADYLPILGAFQDAILTREQOYVEVLDAHPLEW	5640	QY	6721	FLKSNVABEAHFINTKQLKLLARSRQWRSUMSYKSLVNRISPELIRGPPDPSLGVAR	6780
QY	5641	LVRTPTKSSPRQGWSPAYLDRRLKLSPEWGAEEPEFGEAVSEDEYKARLSSVIOE	5700				

Db 6721 FLKMPAAEEAHFINTKQLKFLARSRWQSLMSYKSLVNRISPEILLRQPPSPSLGVAR 6780  
Qy 6781 HLCRDGTGSSSSSSSDNELAPPARAKSLPPSVTHSPLLHPRGFLRPSASLPEEAASE 6840  
Db 6781 HLCRDGTGSSSSSSSDNELAPPARAKSLPPSVTHSPLLHPRGFLRPSASLPEEAASE 6840  
Qy 6841 RSTAPAPPASPEGAGPPAAQGVPHSVIRSLFYHOAGSEPHGALAPGSRHRPARRH 6900  
Db 6841 RSTAPAPPASPEGAGPPAAQGVPHSVIRSLFYHOAGSEPHGALAPGSRHRPARRH 6900  
Qy 6901 LKGGYIAGALPGLREPLMEHRVLEBAAREEATLLAKAPSPETALRLPASGTHLAPGH 6960  
Db 6901 LKGGYIAGALPGLREPLMEHRVLEBAAREEATLLAKAPSPETALRLPASGTHLAPGH 6960  
Qy 6961 SHSLEHDSPTPRPSSACGAQLPSAPSGGAPIRDMGHPQSGKQLPSTGGHPGTAQPE 7020  
Db 6961 SHSLEHDSPTPRPSSACGAQLPSAPSGGAPIRDMGHPQSGKQLPSTGGHPGTAQPE 7020  
Qy 7021 RPSDSWGWOPAPFCHPKQSAQEGCSPPHAPVAPCPGSPFGSCKEAPLVSPSPFLQG 7080  
Db 7021 RPSDSWGWOPAPFCHPKQSAQEGCSPPHAPVAPCPGSPFGSCKEAPLVSPSPFLQG 7080  
Qy 7081 PQAPPAKASPPILDSXWGFCDISLPCRPKPGCSPGSSASQSSQVSLRVGSSQVGT 7140  
Db 7081 PQAPPAKASPPILDSXWGFCDISLPCRPKPGCSPGSSASQSSQVSLRVGSSQVGT 7140  
Qy 7141 EPGSLDAEGWTOAEADLSSTPTLQRPQCATWRKFSLGRGGYAGVAGYGFAGGDA 7200  
Db 7141 EPGSLDAEGWTOAEADLSSTPTLQRPQCATWRKFSLGRGGYAGVAGYGFAGGDA 7200  
Qy 7201 GGMGLQGPMWARIWAVSQSEEEQBEARAESEQEEQAEARAESEPLQVBARPVPEVGR 7260  
Db 7201 GGMGLQGPMWARIWAVSQSEEEQBEARAESEQEEQAEARAESEPLQVBARPVPEVGR 7260  
Qy 7261 PTRSSPPTMEDIGQVSLVQIRLSDGDAEAADTISLDISEVDPAYNLNLSLDYDKYLPF 7320  
Db 7261 PTRSSPPTMEDIGQVSLVQIRLSDGDAEAADTISLDISEVDPAYNLNLSLDYDKYLPF 7320  
Qy 7321 EFMIFRKVPKSAQPEPPSPMAEELAEFFETWPGELPHAGLEITESESDVALLAE 7380  
Db 7321 EFMIFRKVPKSAQPEPPSPMAEELAEFFETWPGELPHAGLEITESESDVALLAE 7380  
Qy 7381 AAVGKRKWSPSLSLHFFPGHPLDPEAPLGLRERVKASVEHISILKGRPEGLKEG 7440  
Db 7381 AAVGKRKWSPSLSLHFFPGHPLDPEAPLGLRERVKASVEHISILKGRPEGLKEG 7440  
Qy 7441 PPRKPGGLASPRLSGLKSWDRAPTFLRLSDTETVVGOSVTLACQVSAQAQAATWSKDG 7500  
Db 7441 PPRKPGGLASPRLSGLKSWDRAPTFLRLSDTETVVGOSVTLACQVSAQAQAATWSKDG 7500  
Qy 7501 APLESSRVLSATLKNFOLLTILVVAEDLGVTCSVSNALGTVTITGVLKRAERSSS 7560  
Db 7501 APLESSRVLSATLKNFOLLTILVVAEDLGVTCSVSNALGTVTITGVLKRAERSSS 7560  
Qy 7561 PCPDIGEVYADGVLLVWKPVESYGVPTVIVQCSLEGGSWTTLASDIPDCCYLTSKLSRG 7620  
Db 7561 PCPDIGEVYADGVLLVWKPVESYGVPTVIVQCSLEGGSWTTLASDIPDCCYLTSKLSRG 7620  
Qy 7621 TTTFTACVSKAGMGPYSPSPQVLLGGPSHLASEESQGSQAOLPSTKTFAFQTQIQ 7680  
Db 7621 TTTFTACVSKAGMGPYSPSPQVLLGGPSHLASEESQGSQAOLPSTKTFAFQTQIQ 7680  
Qy 7681 GRFSVVRQWCKASGRALAAKIIPYHPKDKTAVLEYEALKGLRHPHLAQLHAAYLSPRH 7740  
Db 7681 GRFSVVRQWCKASGRALAAKIIPYHPKDKTAVLEYEALKGLRHPHLAQLHAAYLSPRH 7740  
Qy 7741 LVLLLELCGSPPELLPCLABRASYSESEVKDYLMQMSATQYLNHQLHLDLRSENMIIT 7800  
Db 7741 LVLLLELCGSPPELLPCLABRASYSESEVKDYLMQMSATQYLNHQLHLDLRSENMIIT 7800  
Qy 7801 EYNLLKVDLGNASQISQEKVLPSPDKFDYLETMAPELLEGOCVAPQTDIWAIGVTAFIM 7860  
Db 7801 EYNLLKVDLGNASQISQEKVLPSPDKFDYLETMAPELLEGOCVAPQTDIWAIGVTAFIM 7860

RESULT 2

US-10-080-334-290  
; Sequence 290, Application US/10080334  
; Publication No. US20040002584A1  
; GENERAL INFORMATION:  
; APPLICANT: Pena, Carol E. A.  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Li, Li  
; APPLICANT: Shenoy, Suresh G  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Vernet, Corine A. M.  
; APPLICANT: Malyankar, Uriel M  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Gusev, Vladimir Y  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Furtak, Katarzyna  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Baumgartner, Jason C.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Spaderna, Steven K  
; APPLICANT: Zernusen, Bryan D  
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of  
; FILE REFERENCE: 21402-275  
; CURRENT APPLICATION NUMBER: US/10/080,334  
; CURRENT FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: 60/270,523  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR APPLICATION NUMBER: 60/322,712  
; PRIOR FILING DATE: 2001-09-17  
; PRIOR APPLICATION NUMBER: 60/311,980  
; PRIOR FILING DATE: 2001-08-13  
; PRIOR APPLICATION NUMBER: 60/330,307  
; PRIOR FILING DATE: 2001-10-18  
; PRIOR APPLICATION NUMBER: 60/278,796  
; PRIOR FILING DATE: 2001-03-26  
; PRIOR APPLICATION NUMBER: 60/281,521  
; PRIOR FILING DATE: 2001-04-04  
; PRIOR APPLICATION NUMBER: 60/276,677  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/311,595  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/270,220  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR APPLICATION NUMBER: 60/274,295  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/318,526  
; PRIOR FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/286,548  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: 60/291,765  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/270,797  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: 60/276,400  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/270,810

; PRIOR FILING DATE: 2001-02-23  
 ; NUMBER OF SEQ ID NOS: 388  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 290  
 ; LENGTH: 6620  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-080-334-290

Query Match 77.9%; Score 32134; DB 15; Length 6620;

Best Local Similarity 98.1%; Pred. No. 0;  
Matches 6250; Conservative 15; Mismatches 57; Indels 48; Gaps 9;

QY	1	MDQPFSGAPFLTRPKAFVSVGKDATLSQIVGNPTPQVSWKDDQOPVTAGARFLAQ	60
DB	1	MDQPFSGAPFLTRPKAFVSVGKDATLSQIVGNPTPQVSWKDDQOPVTAGARFLAQ	60
QY	61	DGDLVRLTLDLALGDSQYVCRARNATGEAFAAAGLVQVDAEACACQAPHLRPTSI	120
DB	61	DGDLVRLTLDLALGDSQYVCRARNATGEAFAAAGLVQVDAEACACQAPHLRPTSI	120
QY	121	VREGSEATFCRCVGSPPAPVSWKDGRRLLGEPDGRVVEELGEASALRIRARPRDGG	180
DB	121	VREGSEATFCRCVGSPPAPVSWKDGRRLLGEPDGRVVEELGEASALRIRARPRDGG	180
QY	181	TYEVRAENPLCAASAAALVDSADATASPGTSTAALLAHQORREANRAEGAPASP	240
DB	181	TYEVRAENPLCAASAAALVDSADATASPGTSTAALLAHQORREANRAEGAPASP	240
QY	241	STGTRTCTVTGKCHARLSYVTGPKPETVWKDGQLVTEGRHVVDQAENFVLKILF	300
DB	241	STGTRTCTVTGKCHARLSYVTGPKPETVWKDGQLVTEGRHVVDQAENFVLKILF	300
QY	301	CQSDRGLYCTANLVGQTVSSVLVVRBPAPVFKRLQDLEVRKESATFLCEVPQPS	360
DB	301	CQSDRGLYCTANLVGQTVSSVLVVRBPAPVFKRLQDLEVRKESATFLCEVPQPS	360
QY	361	TEAAWKEETELWASAKYIEEGTERRLTVRNVSADDDAVYICETPEGSRTVAELAVQ	420
DB	361	TEAAWKEETELWASAKYIEEGTERRLTVRNVSADDDAVYICETPEGSRTVAELAVQ	420
QY	421	NLLRKLPRKTAVRVGDVTAMFCVELAVPGVPHLNRQEEVVAAGRVVAISAEGRHTLTIS	480
DB	421	NLLRKLPRKTAVRVGDVTAMFCVELAVPGVPHLNRQEEVVAAGRVVAISAEGRHTLTIS	480
QY	481	QCCLLEDVGOAFMAGDCQSTRFCVSAAPRPPLOPPVDVVKVARMESVILSWSPPHGE	540
DB	481	QCCLLEDVGOAFMAGDCQSTRFCVSAAPRPPLOPPVDVVKVARMESVILSWSPPHGE	540
QY	541	RPVTIDGLVVEKKLGTYYTIRCHEABWATPELTIVADVAEENGFQFRYSALNSFCQSPY	600
DB	541	RPVTIDGLVVEKKLGTYYTIRCHEABWATPELTIVADVAEENGFQFRYSALNSFCQSPY	600
QY	601	LEPGTVHLAPKLAVRTPLKAVQAVEGGEVTFPSVDLTVASAGFWFLDQALKASSVYEIH	660
DB	601	LEPGTVHLAPKLAVRTPLKAVQAVEGGEVTFPSVDLTVASAGFWFLDQALKASSVYEIH	660
QY	661	CDTRHTLTITREVPASLHGALKFVANGIESSIRMEVRAAPGLTANKPPAAAREVLARL	720
DB	661	CDTRHTLTITREVPASLHGALKFVANGIESSIRMEVRAAPGLTANKPPAAAREVLARL	720
QY	721	HEEAQALLAEISDQAAAVTWLKDQRTLSPGPKYEVQASAGRRVLLVRDVARDDAGLYECVS	780
DB	721	HEEAQALLAEISDQAAAVTWLKDQRTLSPGPKYEVQASAGRRVLLVRDVARDDAGLYECVS	780
QY	781	RGRIAYQISVQGLARFLHKDMAGSCVDVAGGPAQFECETSEAHVHVHWYKDGVELGHS	840
DB	781	RGRIAYQISVQGLARFLHKDMAGSCVDVAGGPAQFECETSEAHVHVHWYKDGVELGHS	840
QY	841	GERFLOEDVGTTRHLVAATVTRDEGTYSQVSGEDSVDFLRVSPKVVFAKEQLARRKL	900
DB	841	GERFLOEDVGTTRHLVAATVTRDEGTYSQVSGEDSVDFLRVSPKVVFAKEQLARRKL	900

QY	901	QABAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVCMCEATCTRRLLVVOQAGQADAGEY	960
DB	901	QABAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVCMCEATCTRRLLVVOQAGQADAGEY	960
QY	961	SCEAGGQRLSFHLDVKEPKVFAKQVAHSEVQAEAGANATLSCEVAQAQAEVWYKDGK	1020
DB	961	SCEAGGQRLSFHLDVKEPKVFAKQVAHSEVQAEAGANATLSCEVAQAQAEVWYKDGK	1020
QY	1021	KLSSSLKHVVEAKGCRRLVVOQAKTADGYSCEARGQVSRFLHITTEPKWFAKEQSV	1080
DB	1021	KLSSSLKHVVEAKGCRRLVVOQAKTADGYSCEARGQVSRFLHITTEPKWFAKEQSV	1080
QY	1081	HNEVQAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVMEVKGCTRRLLVPOAGKAD	1140
DB	1081	HNEVQAEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVMEVKGCTRRLLVPOAGKAD	1140
QY	1141	AGEYSCEAGGQVSRFLHITTEPKVFAKEQSVHNEVQAEAGTATMLSCVAQPOTEVTWY	1200
DB	1141	AGEYSCEAGGQVSRFLHITTEPKVFAKEQSVHNEVQAEAGTATMLSCVAQPOTEVTWY	1200
QY	1201	KDGKLSSSSKVMEVKGCTRRLLVVOQKADAGEYSCEAGGQVSRFLHITTEPKVFAK	1260
DB	1201	KDGKLSSSSKVMEVKGCTRRLLVVOQKADAGEYSCEAGGQVSRFLHITTEPKVFAK	1260
QY	1261	EQLVHNEVTEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVRIEAGCMRQLVVOQA	1320
DB	1261	EQLVHNEVTEAGASATLSCEVAQAQTEVTWYKDGKLSSSSKVRIEAGCMRQLVVOQA	1320
QY	1321	GOADAGEVTCBAGGQRLSFHLDVSEPKAVPAKEQLAHRKVOAEAGATATLSCEVAQAQTE	1380
DB	1321	GOADAGEVTCBAGGQRLSFHLDVSEPKAVPAKEQLAHRKVOAEAGATATLSCEVAQAQTE	1380
QY	1381	VTWYKDGKLSSSSKVMEVKGCTRRLLVVOQAQADTGEYSCEAGGQRLSFSLDVAPKV	1440
DB	1381	VTWYKDGKLSSSSKVMEVKGCTRRLLVVOQAQADTGEYSCEAGGQRLSFSLDVAPKV	1440
QY	1441	VFAKEQPVHREVQAGASTTLSCVAQAQTEVTWYKDGKLSFSSKVRMEAVGCTRRLLV	1500
DB	1441	VFAKEQPVHREVQAGASTTLSCVAQAQTEVTWYKDGKLSFSSKVRMEAVGCTRRLLV	1500
QY	1501	VQAGQADAGEYSCEAGSQRSLSFHLVHAEPAKPAKQASREVEQAEAGTATLSCEVAQ	1560
DB	1501	VQAGQADAGEYSCEAGSQRSLSFHLVHAEPAKPAKQASREVEQAEAGTATLSCEVAQ	1560
QY	1561	AQTEVTWYKDGKLSSSSKVMEVKGCTRRLLVVOQAQADAGEYSCKAGQRLSFHLVA	1620
DB	1561	AQTEVTWYKDGKLSSSSKVMEVKGCTRRLLVVOQAQADAGEYSCKAGQRLSFHLVA	1620
QY	1621	BPVVFAKEQPAHREVQAEAGASATLSCEVAQAQTEVTWYKDGKLSFSSSKVRMEAVGCT	1680
DB	1621	BPVVFAKEQPAHREVQAEAGASATLSCEVAQAQTEVTWYKDGKLSFSSSKVRMEAVGCT	1680
QY	1681	RRLVVOQAQADAGEYSCEAGGQRLSFHLVHAELEPOISERPCCRREPLVVKHEHDIILTA	1740
DB	1681	RRLVVOQAQADAGEYSCEAGGQRLSFHLVHAELEPOISERPCCRREPLVVKHEHDIILTA	1740
QY	1741	TLATPSAATVTWYKDGVEIRSRKHETASQGDTHTLTVHGAQVLDLSAISCVRGAEQDF	1800
DB	1741	TLATPSAATVTWYKDGVEIRSRKHETASQGDTHTLTVHGAQVLDLSAISCVRGAEQDF	1800
QY	1801	PVQVEEVAAKFCRLLEPVCGLGTTVLAELSPAACEVWVRGNTQPRVGKQFQWVAEG	1860
DB	1801	PVQVEEVAAKFCRLLEPVCGLGTTVLAELSPAACEVWVRGNTQPRVGKQFQWVAEG	1860
QY	1861	PVRSITVILGAEADAGEYCESDDHTSAQLTIVSVPRVVKFMSGLSTVVAEAGEATFQC	1920
DB	1861	PVRSITVILGAEADAGEYCESDDHTSAQLTIVSVPRVVKFMSGLSTVVAEAGEATFQC	1920
QY	1921	VSPSDVAVVWVRDQALLQPSKFAISQSGASHLTIISDLVLEDAQOITVEAEGSSAA	1980
DB	1921	VSPSDVAVVWVRDQALLQPSKFAISQSGASHLTIISDLVLEDAQOITVEAEGSSAA	1980
QY	1981	LRVREAPVLFKKLEPQTVSEERSVTLELTPWPELKRWRNATAPALAPKKNVEIHAEGA	2040

Db 1981 LRVREAPVLFKKLEPQTVSESSVTLLEVELTREPWELRWRNATALAPGNVEIHAEGA 2040  
Qy RHRLVHNVPADRGFFGCEPDDKTQAKLTVENRQVRLVRLGLOAVAREQGTATMEVOL 2100  
Db RHRLVHNVPADRGFFGCEPDDKTQAKLTVENRQVRLVRLGLOAVAREQGTATMEVOL 2100  
Qy SHADVDSWTRDGLRFQOGPTCHLAVRGPMHTLTLGLRPEDSGLMVFKAGVHTSARLV 2160  
Db SHADVDSWTRDGLRFQOGPTCHLAVRGPMHTLTLGLRPEDSGLMVFKAGVHTSARLV 2160  
Qy VTELVPVSFRPLQDVVITTEKSKVTLCELSRPNDVRLWKDGVRELRAKTMATAAQAACR 2220  
Db VTELVPVSFRPLQDVVITTEKSKVTLCELSRPNDVRLWKDGVRELRAKTMATAAQAACR 2220  
Qy SLTIYRCEFAQGVVCDADHAQSSASVKVQGRYTYLTIYRVLAEADAGEIQFVAENAESR 2280  
Db SLTIYRCEFAQGVVCDADHAQSSASVKVQGRYTYLTIYRVLAEADAGEIQFVAENAESR 2280  
Qy AOLRVKELPVTILVRLPRDKIAMEXHRGVLECOVRSASQAVRPFKGSQELQPGPKYELVSD 2340  
Db AOLRVKELPVTILVRLPRDKIAMEXHRGVLECOVRSASQAVRPFKGSQELQPGPKYELVSD 2340  
Qy GLYRKLIISDVHAEDEDTYTCDAGDVKTSAQFFVEEQSIITIVRGLQDVTVMEPAPAMFEC 2400  
Db GLYRKLIISDVHAEDEDTYTCDAGDVKTSAQFFVEEQSIITIVRGLQDVTVMEPAPAMFEC 2400  
Qy ETSIPSVRPPKWLKGTVLQAGNVGLEBOGTVHRLMLRRTCTMTGPVHFTVCKSRSSA 2460  
Db ETSIPSVRPPKWLKGTVLQAGNVGLEBOGTVHRLMLRRTCTMTGPVHFTVCKSRSSA 2460  
Qy RLIVSDIIPVLTTRPLEPKTGRELOSVLSCDFRPAKPAVOMYKDDTPLSPSEKFKMSLEG 2520  
Db RLIVSDIIPVLTTRPLEPKTGRELOSVLSCDFRPAKPAVOMYKDDTPLSPSEKFKMSLEG 2520  
Qy QMAELRIILRLPADAGVYRCQAGSAHSSTEVTVZAREVTVTGPIQDAEATEEGWASFSCE 2580  
Db QMAELRIILRLPADAGVYRCQAGSAHSSTEVTVZAREVTVTGPIQDAEATEEGWASFSCE 2580  
Qy LSHEDVEVWSLNGMPLVNDSPHETSHKGRHTLVLSIQADAGIVRASLSKYVSTASRL 2640  
Db LSHEDVEVWSLNGMPLVNDSPHETSHKGRHTLVLSIQADAGIVRASLSKYVSTASRL 2640  
Qy EVRVKPVVFLKALDLSAEERGTALQCEVSDPEAHVVRKDXGVLQGFSDKYDFLHTAGT 2700  
Db EVRVKPVVFLKALDLSAEERGTALQCEVSDPEAHVVRKDXGVLQGFSDKYDFLHTAGT 2700  
Qy RGLVHDVSPEDAGLYTCHVGSETRARVVRHDLHVGITKELKTMEVLEGBSCSFECVLS 2760  
Db RGLVHDVSPEDAGLYTCHVGSETRARVVRHDLHVGITKELKTMEVLEGBSCSFECVLS 2760  
Qy HESASDPAMWTVGGKTVGSSRFQATROGRKYILVREAAPSDAGEVVFVSRGLTSKASL 2820  
Db HESASDPAMWTVGGKTVGSSRFQATROGRKYILVREAAPSDAGEVVFVSRGLTSKASL 2820  
Qy IYRERPAALIKPLEQWVAPGSDVBLRCELGRAGTPVHVLKDRKAIKRSQKYDVVVCBGTM 2880  
Db IYRERPAALIKPLEQWVAPGSDVBLRCELGRAGTPVHVLKDRKAIKRSQKYDVVVCBGTM 2880  
Qy AMLVIRGASLKDAGSYTCEVEASKSTASLHVEEKANCFTTELTLNLOVEEKGTAVFTCKTE 2940  
Db AMLVIRGASLKDAGSYTCEVEASKSTASLHVEEKANCFTTELTLNLOVEEKGTAVFTCKTE 2940  
Qy HPAATVTRWKGLELRASGHQPSQEGTLRTISALEKADSDTYTCDIGQAQRAQLLV 3000  
Db HPAATVTRWKGLELRASGHQPSQEGTLRTISALEKADSDTYTCDIGQAQRAQLLV 3000  
Qy QGRRVHIIBDLDDVDVQEGSSATFCRISPANYEPVHNFLDKTPHLANELNEIDAQPGGY 3060  
Db QGRRVHIIBDLDDVDVQEGSSATFCRISPANYEPVHNFLDKTPHLANELNEIDAQPGGY 3060  
Qy HVLTLRQALKDSGTIYFAGQORASALRVTEKPSVFSRELTDTATITEGEDTLTVLCETS 3120

Db 3061 HVLTLRQALKDSGTIYFAGQORASALRVTEKPSVFSRELTDTATITEGEDTLTVLCETS 3120  
Qy TCDIPMCWTKDGTLRGSARCOLSHEGHRAQLLIITGATLQDSGRYKCBAGGACSSSIVRV 3180  
Db TCDIPMCWTKDGTLRGSARCOLSHEGHRAQLLIITGATLQDSGRYKCBAGGACSSSIVRV 3180  
Qy HARPVRFQALKDLEVLLEGGAATLRCVLSSVAAPVKWCYGNVLRPDKYSURQSGAMLE 3240  
Db HARPVRFQALKDLEVLLEGGAATLRCVLSSVAAPVKWCYGNVLRPDKYSURQSGAMLE 3240  
Qy LVVRNLRPDSGRYSCSGFDQTTSATLTVTALPAQFIGLKRNEKATEGATATLRCELSKT 3300  
Db LVVRNLRPDSGRYSCSGFDQTTSATLTVTALPAQFIGLKRNEKATEGATATLRCELSKT 3300  
Qy APVWRKGSSETLRDGDYCLRDQGANCELOIRGLAMVDAABYSCVCGEERTSASLTIRM 3360  
Db APVWRKGSSETLRDGDYCLRDQGANCELOIRGLAMVDAABYSCVCGEERTSASLTIRM 3360  
Qy PAHFIGRLRHQSIEGATATLRCELSKAAPVWRKGRESLRDGDHRSRAQDGAVALCQIC 3420  
Db PAHFIGRLRHQSIEGATATLRCELSKAAPVWRKGRESLRDGDHRSRAQDGAVALCQIC 3420  
Qy GLAVADAGYSYCVGGBERTSATLTVKALPAKTEGLRNEEAVEGATAMLWCELSKVAPVE 3480  
Db GLAVADAGYSYCVGGBERTSATLTVKALPAKTEGLRNEEAVEGATAMLWCELSKVAPVE 3480  
Qy WRKGPENLRDGDYILRQEGTRCELOICGLAMADAGEYLCVCGQERTSATLTIRALPARF 3540  
Db WRKGPENLRDGDYILRQEGTRCELOICGLAMADAGEYLCVCGQERTSATLTIRALPARF 3540  
Qy IEDVKNQEARBEGATAVLOCELNSAAPVWRKGSSETLRDGDYSLRDQGTGKCBQIRGLAM 3600  
Db IEDVKNQEARBEGATAVLOCELNSAAPVWRKGSSETLRDGDYSLRDQGTGKCBQIRGLAM 3600  
Qy ADTGEYSYCVGGBERTSATLTVRALPIKFTEGRLNEEATEGATAVLCELSKVAPVEMWK 3660  
Db ADTGEYSYCVGGBERTSATLTVRALPIKFTEGRLNEEATEGATAVLCELSKVAPVEMWK 3660  
Qy HETLRDGDHRSLRDQDARCELOIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPSKIEGL 3720  
Db HETLRDGDHRSLRDQDARCELOIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPSKIEGL 3720  
Qy RNEEATEGDTATLWCELSKAAAPVWRKGHETLRDGDHRSLRDQDSRCELOIRGLAVVDAG 3780  
Db RNEEATEGDTATLWCELSKAAAPVWRKGHETLRDGDHRSLRDQDSRCELOIRGLAVVDAG 3780  
Qy EYSCVCGQERTSATLTVRALPARFIEDVKNQEARBEGATAVLOCELNSAAPVWRKGSSETL 3840  
Db EYSCVCGQERTSATLTVRALPARFIEDVKNQEARBEGATAVLOCELNSAAPVWRKGSSETL 3840  
Qy RGGDRYSLRDQDTRCBLQIHGLSVADTGYSCVCGQERTSATLTVRAPQPVFREPQLSLQ 3900  
Db RGGDRYSLRDQDTRCBLQIHGLSVADTGYSCVCGQERTSATLTVRAPQPVFREPQLSLQ 3900  
Qy AEEGSTATLOCELSSEPTATVMSKGLQIQANGREPRLOGCTABEVLVLODLOREDTGEYT 3960  
Db AEEGSTATLOCELSSEPTATVMSKGLQIQANGREPRLOGCTABEVLVLODLOREDTGEYT 3960  
Qy CTCGSQATSATLTVTAAAPVRFLELQHQVEDEGGTAHLCELSRAGASVWRKGSQOLFP 4020  
Db CTCGSQATSATLTVTAAAPVRFLELQHQVEDEGGTAHLCELSRAGASVWRKGSQOLFP 4020  
Qy CAKYQVQDGAALVELVRGVEQDAGDYTCDTGHTOSMASLSVRVPRPKFKTELQSLQE 4080  
Db CAKYQVQDGAALVELVRGVEQDAGDYTCDTGHTOSMASLSVRVPRPKFKTELQSLQE 4080  
Qy TGDARLCCOLSDAESGAVVQWLKEGVHLAGPKYEMRSQGATRELLIHLQEAKTGEYA 4140  
Db TGDARLCCOLSDAESGAVVQWLKEGVHLAGPKYEMRSQGATRELLIHLQEAKTGEYA 4140  
Qy CVTGGQKTAASLRVTEPEVTIVRGLVDAEVTADEDFEFCSEVSRAGATGVQWCLQGLPLQ 4200  
Db CVTGGQKTAASLRVTEPEVTIVRGLVDAEVTADEDFEFCSEVSRAGATGVQWCLQGLPLQ 4200



QY	4201	SNEVTEAVRDRGR	IHTLRLKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVITLPLEQDVQ	4260
DB	4201	SNEVTEAVRDRGR	IHTLRLKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVITLPLEQDVQ	4260
QY	4261	LSEGDASFQCLRS	SPASQEARWALGGVPLQANEMNDITVBOGTLHLTLHKVTFLEDAQT	4320
DB	4261	LSEGDASFQCLRS	SPASQEARWALGGVPLQANEMNDITVBOGTLHLTLHKVTFLEDAQT	4320
QY	4321	VSPHVTGTCSS	EAQLKVTAKNTVVRGLENVVEALEGEALFEQOLSQPEVAATHWLLDDEPV	4380
DB	4321	VSPHVTGTCSS	EAQLKVTAKNTVVRGLENVVEALEGEALFEQOLSQPEVAATHWLLDDEPV	4380
QY	4381	RTSENAEVVFFENGLRH	LLKLNLRPQDSCRVTFLAGDMVTSAFITVRGMRLEILEPLKN	4440
DB	4381	RTSENAEVVFFENGLRH	LLKLNLRPQDSCRVTFLAGDMVTSAFITVRGMRLEILEPLKN	4440
QY	4441	AAVRAGAQAARTCT	ILSEAVPVGEASWYINGAAVOPDDSDMTVTADGSHQALLRSAPHH	4500
DB	4441	AAVRAGAQAARTCT	ILSEAVPVGEASWYINGAAVOPDDSDMTVTADGSHQALLRSAPHH	4500
QY	4501	AGEVTFACRDVAVASARL	TVLGLPDPPEDAEVVAHSSHVTLSWAAPMSDGGGLCGYRVE	4560
DB	4501	AGEVTFACRDVAVASARL	TVLGLPDPPEDAEVVAHSSHVTLSWAAPMSDGGGLCGYRVE	4560
QY	4561	VKEGATGQWRLCH	ELVPGPECVVDGLAPGETYRFRVAAGVGVGAGEPVHLPQTVRLAEP	4620
DB	4561	VKEGATGQWRLCH	ELVPGPECVVDGLAPGETYRFRVAAGVGVGAGEPVHLPQTVRLAEP	4620
QY	4621	KPVPPPSAPBSROVA	AGEDVSLELVVAZAGEVIMHKGMRLOPGGRFEFVSGRQOML	4680
DB	4621	KPVPPPSAPBSROVA	AGEDVSLELVVAZAGEVIMHKGMRLOPGGRFEFVSGRQOML	4680
QY	4681	VIKGFTABDQGEYH	CGLAQGSICPAAATFQVALSPASVDEAPQPSLPEAAQEGDLHLW	4740
DB	4681	VIKGFTABDQGEYH	CGLAQGSICPAAATFQVALSPASVDEAPQPSLPEAAQEGDLHLW	4740
QY	4741	EALARKRMSREPTL	DSISELPEEDGRSQRLPQAEBAVADLSEGYSTADELARTGDADL	4800
DB	4741	EALARKRMSREPTL	DSISELPEEDGRSQRLPQAEBAVADLSEGYSTADELARTGDADL	4800
QY	4801	SHTSSDDES	RAGTSLVTYLLKACRGCTSPILASKVGAAPAAPSVKPQOQOEPLAARPPPLG	4860
DB	4801	SHTSSDDES	RAGTSLVTYLLKACRGCTSPILASKVGAAPAAPSVKPQOQOEPLAARPPPLG	4860
QY	4861	DLSTKDLDGDP	SMDXAAVKIQAAFKGYKVRKEMKQOQEGMFSTFGTEAVQGDALRECV	4920
DB	4861	DLSTKDLDGDP	SMDXAAVKIQAAFKGYKVRKEMKQOQEGMFSTFGTEAVQGDALRECV	4920
QY	4921	VASRADVRARWL	KDGVELTDGRHHHIQOLGDGTCSLLIAGLDRADAGCYTCQVSNKFGOV	4980
DB	4921	VASRADVRARWL	KDGVELTDGRHHHIQOLGDGTCSLLIAGLDRADAGCYTCQVSNKFGOV	4980
QY	4981	THSAVTVVSGSE	SAESSSGGELDDAFRAARLRLFRTKSPAESVDEELFLSADEGPA	5040
DB	4981	THSAVTVVSGSE	SAESSSGGELDDAFRAARLRLFRTKSPAESVDEELFLSADEGPA	5040
QY	5041	EPEEPADQWYRE	DEHFCIRFEALTEARQAVTRFOEMPATLIGVIEIKLVEOGPRRVE	5100
DB	5041	EPEEPADQWYRE	DEHFCIRFEALTEARQAVTRFOEMPATLIGVIEIKLVEOGPRRVE	5100
QY	5101	CISKETPAVVP	PPPLSLTSDAAPVFLTELQNEQVODGPVPSFCVVTGQMPSPVRWF	5160
DB	5101	CISKETPAVVP	PPPLSLTSDAAPVFLTELQNEQVODGPVPSFCVVTGQMPSPVRWF	5160
QY	5161	KDGLLEDDHYM	INEDQCGHQHIIITAVVPADMGVYRCLAENSMGVSSTKAELRVDLTS	5220
DB	5161	KDGLLEDDHYM	INEDQCGHQHIIITAVVPADMGVYRCLAENSMGVSSTKAELRVDLTS	5220
QY	5221	TDYDTAADAT	ESSSYFSAQYLSRREOEGTESTTDEGQLPQVVEELRDQLVAFGTBLAKF	5280
DB	5221	TDYDTAADAT	ESSSYFSAQYLSRREOEGTESTTDEGQLPQVVEELRDQLVAFGTBLAKF	5280
QY	5281	OLKVKGY	PAPRLYWKDGOPLTASAHIRMTGKKILHTLEIISVTRSDSGQAAVYISNMG	5340
DB	5281	OLKVKGY	PAPRLYWKDGOPLTASAHIRMTGKKILHTLEIISVTRSDSGQAAVYISNMG	5340
QY	5341	AAYSSARLLV	RGDPPEEKFPASDVHEQLVPPRMLERFTPKVKYKGGSSITFSVKVEGRVP	5400
DB	5341	AAYSSARLLV	RGDPPEEKFPASDVHEQLVPPRMLERFTPKVKYKGGSSITFSVKVEGRVP	5400
QY	5401	TVHMLRE	EAERGVLMIGDPTPGYTVVASSAQOHSVLVDVGRHQGTTCIASNAAGQALC	5460
DB	5401	TVHMLRE	EAERGVLMIGDPTPGYTVVASSAQOHSVLVDVGRHQGTTCIASNAAGQALC	5460
QY	5461	SASLHVSG	LKPKVEOEKVKREALISTFLOGTTOAISAOGLETASPADLGGQKEEPLAAKE	5520
DB	5461	SASLHVSG	LKPKVEOEKVKREALISTFLOGTTOAISAOGLETASPADLGGQKEEPLAAKE	5520
QY	5521	ALGHLSL	AEVGTETEFLOKLTQOITEMVSAKITQAKLVPGGSDSDSKTSPASPRHGSR	5580
DB	5521	ALGHLSL	AEVGTETEFLOKLTQOITEMVSAKITQAKLVPGGSDSDSKTSPASPRHGSR	5580
QY	5581	PSSSIQESS	SESDGDARGEIYVVTADYLPICAEODAITLREGQVVEVLDAHPPLRW	5640
DB	5581	PSSSIQESS	SESDGDARGEIYVVTADYLPICAEODAITLREGQVVEVLDAHPPLRW	5640
QY	5641	LVRTXPT	KSSPSRGWVSPAYLDRRLKLSPEWGAAAEPEFPGEAVERSEYKARLSSVIOE	5700
DB	5641	LVRTXPT	KSSPSRGWVSPAYLDRRLKLSPEWGAAAEPEFPGEAVERSEYKARLSSVIOE	5700
QY	5701	LSSSOA	QFVEELOFQSHHLOHLERCCHVPIAVAGOKAVIPRNVRDITGRHSSFLQELQ	5760
DB	5701	LSSSOA	QFVEELOFQSHHLOHLERCCHVPIAVAGOKAVIPRNVRDITGRHSSFLQELQ	5760
QY	5761	CTDDDDV	AMCFIKNOAFAFEQYLEFLVGRVQAESVVTSTAIQEFYKKAEEALLAGDPSQ	5820
DB	5761	CTDDDDV	AMCFIKNOAFAFEQYLEFLVGRVQAESVVTSTAIQEFYKKAEEALLAGDPSQ	5820
QY	5821	PPPLCHV	LEQPVVERVQYQALLKELIRNKARNQNCALLQOAVYVSALPQRAENKLHV	5880
DB	5821	PPPLCHV	LEQPVVERVQYQALLKELIRNKARNQNCALLQOAVYVSALPQRAENKLHV	5880
QY	5881	SLMENY	PGTALBGPTRQGHFIVMEGAPGARMKPKGNHVRVFLFRNHLVICKRRDSRT	5940
DB	5881	SLMENY	PGTALBGPTRQGHFIVMEGAPGARMKPKGNHVRVFLFRNHLVICKRRDSRT	5940
QY	5941	DTVSVY	FRNMKLSIDLNDQVGGDDRAFEVWQEREDSVRYKILLOARTAIKSSWVKEIC	6000
DB	5941	DTVSVY	FRNMKLSIDLNDQVGGDDRAFEVWQEREDSVRYKILLOARTAIKSSWVKEIC	6000
QY	6001	GIQOQL	ALPVNRPPDFEELADCTAELGETVKLACRVGTGTPKPIVSWYKDGKAVQVDPHH	6060
DB	6001	GIQOQL	ALPVNRPPDFEELADCTAELGETVKLACRVGTGTPKPIVSWYKDGKAVQVDPHH	6060
QY	6061	ILIEDPD	GCALILDSLTGVDGQVCMFAAAGNCSTGLKILVQVPPRFVNVKVRASPFV	6120
DB	6061	ILIEDPD	GCALILDSLTGVDGQVCMFAAAGNCSTGLKILVQVPPRFVNVKVRASPFV	6120
QY	6121	EGEDAQ	FTCTTEGAPYPOIRWKDQALLTTGNKFTLSEPRSGLLVIVRAASKEDGLY	6180
DB	6121	EGEDAQ	FTCTTEGAPYPOIRWKDQALLTTGNKFTLSEPRSGLLVIVRAASKEDGLY	6180
QY	6181	ECELVNL	GSARASAEILRIQSPMLQAOEQCHREQLVAIVEDTTLER	6232
DB	6181	ECELVNL	GSARASAEILRIQSPMLQAOEQCHREQLVAIVEDTTLER	6232
QY	6233	SVLXEL	GPAPGPGSTGLTGPGCPRG	6271
DB	6241	TVVKS	PRGQRR-SFSKSPSRSPSCSASPRPGLLADPLLYLPAGQPRPEABPGQKPV	6299
QY	6272	-PVTGT	SEAPVPRV-----POLLHEGPE-----QEPALARAQEWTPIRMGA	6317
DB	6300	VPTLVVTE	AEAAHSPALPCLSGPQPKVVEETIEVRVKMGPGQVSPTE--VPRSSSGH	6357
QY	6318	AW--PG	ATG 6325	



Db 6358 LFTLPGATPG 6367  
 RESULT 3  
 US-10-408-765A-2291  
 ; Sequence 2291, Application US/10408765A  
 ; Publication No. US20040101874A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ghosh, Soumitra S.  
 ; APPLICANT: Fahy, Eoin D.  
 ; APPLICANT: Zhang, Bing  
 ; APPLICANT: Gibson, Bradford W.  
 ; APPLICANT: Taylor, Steven W.  
 ; APPLICANT: Glenn, Gary M.  
 ; APPLICANT: Warnock, Dale E.  
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
 ; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
 ; FILE REFERENCE: 660088.465  
 ; CURRENT APPLICATION NUMBER: US/10/408,765A  
 ; CURRENT FILING DATE: 2003-04-04  
 ; NUMBER OF SEQ ID NOS: 3077  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2291:  
 ; LENGTH: 6620  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-408-765A-2291  
 Query Match 77.9%; Score 32134; DB 16; Length 6620;  
 Best Local Similarity 98.1%; Pred. No. 0;  
 Matches 6250; Conservative 15; Mismatches 57; Indels 48; Gaps 9;  
 QY 1 MDQPFSGAPRLTPKAFVSVGKDATLSCOIVGNPTPOYSWEKDQOPVTAGARFLAQ 60  
 Db 1 MDQPFSGAPRLTPKAFVSVGKDATLSCOIVGNPTPOYSWEKDQOPVTAGARFLAQ 60  
 QY 61 DGLYLRTLLDALGDSGGYVCRARNAICEAPFAAVGLQVDAEAAEQAPFLRLPTSIR 120  
 Db 61 DGLYLRTLLDALGDSGGYVCRARNAICEAPFAAVGLQVDAEAAEQAPFLRLPTSIR 120  
 QY 121 VREGSATRCRVGSGPRPAVSWSKDGRRLGEPDGPVRVVEELGEASALIRAAARPDGG 180  
 Db 121 VREGSATRCRVGSGPRPAVSWSKDGRRLGEPDGPVRVVEELGEASALIRAAARPDGG 180  
 QY 181 TYEVRAENPLGAASAAALVDSDAADTASRPGTSTAALLAHLQRRREARAEAGAPASPP 240  
 Db 181 TYEVRAENPLGAASAAALVDSDAADTASRPGTSTAALLAHLQRRREARAEAGAPASPP 240  
 QY 241 STGTRCTVTGEGHARLSCYVTGEPKPTVWKKGQOLVTGEGRRHVVDAGQENFVLKILF 300  
 Db 241 STGTRCTVTGEGHARLSCYVTGEPKPTVWKKGQOLVTGEGRRHVVDAGQENFVLKILF 300  
 QY 301 CKGSDRLGYTCTASNLVGQTYSSVLVVRPAVPFKKRLQDLVREKESATFLCEVPQPS 360  
 Db 301 CKGSDRLGYTCTASNLVGQTYSSVLVVRPAVPFKKRLQDLVREKESATFLCEVPQPS 360  
 QY 361 TEAAWFKETRLWASAKYGIIEBEGTERRLTVERNVSADDDAVYICETPEGSRTVAELAVQG 420  
 Db 361 TEAAWFKETRLWASAKYGIIEBEGTERRLTVERNVSADDDAVYICETPEGSRTVAELAVQG 420  
 QY 421 NLRLKLPRTAVRGDTAMFCVELAVPGPVHWRNQEVEVAGGRVAISAGETHRLTIS 480  
 Db 421 NLRLKLPRTAVRGDTAMFCVELAVPGPVHWRNQEVEVAGGRVAISAGETHRLTIS 480  
 QY 481 QCCLVDVGQVAFWAGDCQSTFCVSAPRKPPLQPPVDPVVKARMESSVILSWSPPHGE 540  
 Db 481 QCCLVDVGQVAFWAGDCQSTFCVSAPRKPPLQPPVDPVVKARMESSVILSWSPPHGE 540  
 QY 541 RPVTIDGYLVKKLGTWTWIRCHEAWVATBELTVADVAEEGNQFPFRVSNALNFGQSPY 600  
 Db 541 RPVTIDGYLVKKLGTWTWIRCHEAWVATBELTVADVAEEGNQFPFRVSNALNFGQSPY 600

Db 1681 RRLVVCQAGADAGEYSCEAGQRLSFLHVAEPEQISERPCREPLVWKHEHDIILTA 1740  
QY 1741 TLATPSAATVWLKDGVEIRRSKRHETASQGDTHTLVHGAQVLDLSAISYCRVGAEGQDF 1800  
Db 1741 TLATPSAATVWLKDGVEIRRSKRHETASQGDTHTLVHGAQVLDLSAISYCRVGAEGQDF 1800  
QY 1801 PVQVEEVAAKTCLRLLEPVCGBLGGTTLTACBLSACAEVVMRCNTQPRVKGKPFQWVAEG 1860  
Db 1801 PVQVEEVAAKTCLRLLEPVCGBLGGTTLTACBLSACAEVVMRCNTQPRVKGKPFQWVAEG 1860  
QY 1861 PVRSLTVLGLRAEDAGEVVCSSRDHDSQAQTVSVPRVWKFMSGLSTVABEGGEATFQC 1920  
Db 1861 PVRSLTVLGLRAEDAGEVVCSSRDHDSQAQTVSVPRVWKFMSGLSTVABEGGEATFQC 1920  
QY 1921 VVSPSDVAVWFRDGLLOPSEKFAISQSGASHSLTISDLVLEDAGQITVVEAEGASSAA 1980  
Db 1921 VVSPSDVAVWFRDGLLOPSEKFAISQSGASHSLTISDLVLEDAGQITVVEAEGASSAA 1980  
QY 1981 LRVREAPVLFKKLEPQTVBERSVTLLEVELTRPMPBLRWTRNATALAPGNVEIHAEGA 2040  
Db 1981 LRVREAPVLFKKLEPQTVBERSVTLLEVELTRPMPBLRWTRNATALAPGNVEIHAEGA 2040  
QY 2041 RHRLVHNVGFDAGRGFFGCEPDDKTQAKLTVMRQVRLVRGLQAVEAREQGTATMEVQL 2100  
Db 2041 RHRLVHNVGFDAGRGFFGCEPDDKTQAKLTVMRQVRLVRGLQAVEAREQGTATMEVQL 2100  
QY 2101 SHADVDSWTRDGLRFQGGPTCHLAVRGPMTHTLTLGLRPEDSGLMVFKAEGVHTSARLV 2160  
Db 2101 SHADVDSWTRDGLRFQGGPTCHLAVRGPMTHTLTLGLRPEDSGLMVFKAEGVHTSARLV 2160  
QY 2161 VTELVPFSRPLQDVWTEKEKVTLECELSRPNDVWELKDGVELRAGKTMALAAQACR 2220  
Db 2161 VTELVPFSRPLQDVWTEKEKVTLECELSRPNDVWELKDGVELRAGKTMALAAQACR 2220  
QY 2221 SLTYRCEPADQGVYVCDADHAQSSASVKVQGRYTTLYIRRVLAEDAGEIQFVAENAESR 2280  
Db 2221 SLTYRCEPADQGVYVCDADHAQSSASVKVQGRYTTLYIRRVLAEDAGEIQFVAENAESR 2280  
QY 2281 AOLRVKELPVTLRPLRDKTAMEKHGVLCEQVSRASAOVRPFKGSQELQPGKYELVSD 2340  
Db 2281 AOLRVKELPVTLRPLRDKTAMEKHGVLCEQVSRASAOVRPFKGSQELQPGKYELVSD 2340  
QY 2341 GLYKRLIISDVHAEDEDTYCDAGDVKTSAQFFVVEQSITIVRGLQDVWMEPAPAFEC 2400  
Db 2341 GLYKRLIISDVHAEDEDTYCDAGDVKTSAQFFVVEQSITIVRGLQDVWMEPAPAFEC 2400  
QY 2401 ETSIPSVRPPKWLGLKTVLQAGNVGLQESGTVHRLMLRSTMTGCPVHFTVKGSRSSA 2460  
Db 2401 ETSIPSVRPPKWLGLKTVLQAGNVGLQESGTVHRLMLRSTMTGCPVHFTVKGSRSSA 2460  
QY 2461 RLVSVDIPVLTLPLEPKTGRELQSVVLSCDFPAPKAVQWYKDDTPLSPSEKFKVLSLEG 2520  
Db 2461 RLVSVDIPVLTLPLEPKTGRELQSVVLSCDFPAPKAVQWYKDDTPLSPSEKFKVLSLEG 2520  
QY 2521 QMAELRLILMPADAGVYRQAGSAHSSTVTEAREBVTVTGLOABATEBGWASFCE 2580  
Db 2521 QMAELRLILMPADAGVYRQAGSAHSSTVTEAREBVTVTGLOABATEBGWASFCE 2580  
QY 2581 LSHDEDEVEVSLNGMPLYNDSFHEISHKGRHRTLVLKSIORADAGIVRASSLKVSTSARL 2640  
Db 2581 LSHDEDEVEVSLNGMPLYNDSFHEISHKGRHRTLVLKSIORADAGIVRASSLKVSTSARL 2640  
QY 2641 EVRVKPVVFLKALDDLSAEBRGTLALQCEVSDPEAHVVMKQGVQIGPSDKYDFLHTAGT 2700  
Db 2641 EVRVKPVVFLKALDDLSAEBRGTLALQCEVSDPEAHVVMKQGVQIGPSDKYDFLHTAGT 2700  
QY 2701 RGLVVHDSPEADAGLYTCHVGSSETRARVRVHDLHVGIITKRLTMEVLGESCSFECVLVS 2760  
Db 2701 RGLVVHDSPEADAGLYTCHVGSSETRARVRVHDLHVGIITKRLTMEVLGESCSFECVLVS 2760  
QY 2761 HESASDPAMWTVGGTKTQVSSSRFOATROGRKYILVWREAPSDAGEVVFVRGLTTSKASL 2820

Db 2761 HESASDPAMWTVGGTKTQVSSSRFOATROGRKYILVWREAPSDAGEVVFVRGLTTSKASL 2820  
QY 2821 IYRERPAALIKPLEDQWVAPGEDVELRCELSPAGTVPVHMLKDKRAIKRSQKYDVVCEGTM 2880  
Db 2821 IYRERPAALIKPLEDQWVAPGEDVELRCELSPAGTVPVHMLKDKRAIKRSQKYDVVCEGTM 2880  
QY 2881 AMLVIRGASLKDAGBYTCEVEASKSTASLHVEEKANCFTBELTNLOVEBKGTAVFTCKTE 2940  
Db 2881 AMLVIRGASLKDAGBYTCEVEASKSTASLHVEEKANCFTBELTNLOVEBKGTAVFTCKTE 2940  
QY 2941 HPAATVTVWRKGLLELURASGKHQPSOEGTLTLTISALEKADSDTTCIDIGOAQRAQLLV 3000  
Db 2941 HPAATVTVWRKGLLELURASGKHQPSOEGTLTLTISALEKADSDTTCIDIGOAQRAQLLV 3000  
QY 3001 QGRRVHIIIEDLVVDVQEGSSATFRCRISPANYEPVHWFMDKTLPLHANELNEIDAQPGY 3060  
Db 3001 QGRRVHIIIEDLVVDVQEGSSATFRCRISPANYEPVHWFMDKTLPLHANELNEIDAQPGY 3060  
QY 3061 HVLTLRQALKDQSGTIYFEAGDQASAAALRVTEKPSVPSRELTDATITTEGEDTLVCETS 3120  
Db 3061 HVLTLRQALKDQSGTIYFEAGDQASAAALRVTEKPSVPSRELTDATITTEGEDTLVCETS 3120  
QY 3121 TCDIPMCWTQDKGKTLRGSARCOLSHEGHRAQLLITGATLQDSGRYKCEAGCACSSSVRV 3180  
Db 3121 TCDIPMCWTQDKGKTLRGSARCOLSHEGHRAQLLITGATLQDSGRYKCEAGCACSSSVRV 3180  
QY 3181 HARPVRFQALKDLEVLGGAAATLRCVLSSVAAAPVKWCYGNVLRPGDKYSLRQEGAMLE 3240  
Db 3181 HARPVRFQALKDLEVLGGAAATLRCVLSSVAAAPVKWCYGNVLRPGDKYSLRQEGAMLE 3240  
QY 3241 LVVRNLRQDQSGRYSCSPGDDTTSATLTVTALPAQFICKLENKEATGATATLCELSKT 3300  
Db 3241 LVVRNLRQDQSGRYSCSPGDDTTSATLTVTALPAQFICKLENKEATGATATLCELSKT 3300  
QY 3301 APVEWRKSETLRQDGRVCLRDQAMCELOIRGLAMVDAABYSVCVCEERTSASLTTRPM 3360  
Db 3301 APVEWRKSETLRQDGRVCLRDQAMCELOIRGLAMVDAABYSVCVCEERTSASLTTRPM 3360  
QY 3361 PAHFTGRIRHOSIEGATATLRCELSKAAPVEWRKGRSLRQDGRHSLRQDGAVCLEQIC 3420  
Db 3361 PAHFTGRIRHOSIEGATATLRCELSKAAPVEWRKGRSLRQDGRHSLRQDGAVCLEQIC 3420  
QY 3421 GLAVADAGEYSCVCEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLCESKVAPVE 3480  
Db 3421 GLAVADAGEYSCVCEERTSATLTVKALPAKFTTEGLRNEEAVEGATAMLCESKVAPVE 3480  
QY 3481 WRKGPENLRDGRYTLRQEGTRCELOICGLAMADAGEVLCVCGOERTSATLITREALPARF 3540  
Db 3481 WRKGPENLRDGRYTLRQEGTRCELOICGLAMADAGEVLCVCGOERTSATLITREALPARF 3540  
QY 3541 IEDVKNQOAREGATAVLQCELSNAPVEWRKGSSETLRDGRYSLRQDGTKELOIRGLAM 3600  
Db 3541 IEDVKNQOAREGATAVLQCELSNAPVEWRKGSSETLRDGRYSLRQDGTKELOIRGLAM 3600  
QY 3601 ADTGEYSVCVCEERTSATLTVRALPIKFTTEGLRNEEAVEGATAMLCESKVAPVEWVK 3660  
Db 3601 ADTGEYSVCVCEERTSATLTVRALPIKFTTEGLRNEEAVEGATAMLCESKVAPVEWVK 3660  
QY 3661 HETLRDGRHSLRQDGRARCELQIRGLVAEDAGEVLCMCKERTSAMLTVRAMPSKFIEGL 3720  
Db 3661 HETLRDGRHSLRQDGRARCELQIRGLVAEDAGEVLCMCKERTSAMLTVRAMPSKFIEGL 3720  
QY 3721 RNEEATEGDTATLCELSSKAAAPVEWRKGHETLRDGRHSLRQDGRSCELQIRGLAVVDAG 3780  
Db 3721 RNEEATEGDTATLCELSSKAAAPVEWRKGHETLRDGRHSLRQDGRSCELQIRGLAVVDAG 3780  
QY 3781 EYSCVCGOERTSATLTVRALPARFIEDVKNQOAREGATAVLQCELSKAAAPVEWRKGSSETL 3840  
Db 3781 EYSCVCGOERTSATLTVRALPARFIEDVKNQOAREGATAVLQCELSKAAAPVEWRKGSSETL 3840  
QY 3841 RGGDRYSLRQDGTKELOICQHGLSVADTGEYSVCVCGOERTSATLTVRAPQVFPREPLOSQ 3900  
Db 3841 RGGDRYSLRQDGTKELOICQHGLSVADTGEYSVCVCGOERTSATLTVRAPQVFPREPLOSQ 3900

QY	3901	ABEGTATLQCBLSPTATVWWSKGLQLOQANGREPREPLOGCTABELVLQDLQREBDTGEYT	3960
Db	3901	ABEGTATLQCBLSPTATVWWSKGLQLOQANGREPREPLOGCTABELVLQDLQREBDTGEYT	3960
QY	3961	CTCGSOATSATLTVTAAPVRFLRELQHOEVDGEGTAHLCCLBSRAGASVEMRKSLQLFPP	4020
Db	3961	CTCGSOATSATLTVTAAPVRFLRELQHOEVDGEGTAHLCCLBSRAGASVEMRKSLQLFPP	4020
QY	4021	CAKYQWQDGAALLLVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKPKTRLOSLEQE	4080
Db	4021	CAKYQWQDGAALLLVRGVEQEDAGDYTCDTGHTQSMASLSVRVPRPKPKTRLOSLEQE	4080
QY	4081	TGDIARLCCOLSDAESGAVQWMLKEGVELHAGPKYEMESQCATRELLIHQLEAKDTGEYA	4140
Db	4081	TGDIARLCCOLSDAESGAVQWMLKEGVELHAGPKYEMESQCATRELLIHQLEAKDTGEYA	4140
QY	4141	CVTGQOKTAASLRVTEPEVTIVRGLVDAEVTADEDEVFSCVSRAGATGVQWCLQGLPLQ	4200
Db	4141	CVTGQOKTAASLRVTEPEVTIVRGLVDAEVTADEDEVFSCVSRAGATGVQWCLQGLPLQ	4200
QY	4201	SNEVTEAVRDRGRIHTLRLKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVTILPELQDVQ	4260
Db	4201	SNEVTEAVRDRGRIHTLRLKGVTPEDAGTVSFHLGNHASSAQLTVRAPEVTILPELQDVQ	4260
QY	4261	LSEGQDASFQCLSRASQOEARWALGGVPLQANEMNDITVBQGTLLHLTLHKVTLDEAGT	4320
Db	4261	LSEGQDASFQCLSRASQOEARWALGGVPLQANEMNDITVBQGTLLHLTLHKVTLDEAGT	4320
QY	4321	VSFHVGTCSSEALQKVTAKNTVVRGLENVLEALEGGEALFECQLSQPEVAHAHTWLLDDEPV	4380
Db	4321	VSFHVGTCSSEALQKVTAKNTVVRGLENVLEALEGGEALFECQLSQPEVAHAHTWLLDDEPV	4380
QY	4381	RTSENAEVVFENGRLHLLLNKLRPQDSRCVTFFLAGDMVTSAPLTVRGWRLLELLEPLKN	4440
Db	4381	RTSENAEVVFENGRLHLLLNKLRPQDSRCVTFFLAGDMVTSAPLTVRGWRLLELLEPLKN	4440
QY	4441	AAVRAGAQAFTCTLSEAVPVGEASWYINGAAVQPDSDMTVTADGSHQALLLSAQPHH	4500
Db	4441	AAVRAGAQAFTCTLSEAVPVGEASWYINGAAVQPDSDMTVTADGSHQALLLSAQPHH	4500
QY	4501	AGEVTFACRDVAVASARLTVILGPDPPEDAENVVAHSSHTVTLSWAAPMSDGGGGJCGYRVE	4560
Db	4501	AGEVTFACRDVAVASARLTVILGPDPPEDAENVVAHSSHTVTLSWAAPMSDGGGGJCGYRVE	4560
QY	4561	VKEGATGWRRLCHELVPGPECVVDGLAPGETYRFRVAAGVPGVAGEPVHLPTQTVRLAEP	4620
Db	4561	VKEGATGWRRLCHELVPGPECVVDGLAPGETYRFRVAAGVPGVAGEPVHLPTQTVRLAEP	4620
QY	4621	KVPVQPSAPESRQVAAAGEDVSLELVVAEAGEVIVHKGMBRIQFGRFVVSQGRQML	4680
Db	4621	KVPVQPSAPESRQVAAAGEDVSLELVVAEAGEVIVHKGMBRIQFGRFVVSQGRQML	4680
QY	4681	VTKGFTAEDQGEYHCGLAQSGICPAAATFOVALSPASVDEAPQPSLPEAAQEGDLHLW	4740
Db	4681	VTKGFTAEDQGEYHCGLAQSGICPAAATFOVALSPASVDEAPQPSLPEAAQEGDLHLW	4740
QY	4741	EALAKRMSREPTLDSISLPEEDGRSQRILPQBAEEVAPDLSEGYSTADELARTGDADL	4800
Db	4741	EALAKRMSREPTLDSISLPEEDGRSQRILPQBAEEVAPDLSEGYSTADELARTGDADL	4800
QY	4801	SHTSSDDDESRACTPSLNTVYLKAGRPGTSPLASKVGAPAAAPSVKPOQOOEPLAARPLG	4860
Db	4801	SHTSSDDDESRACTPSLNTVYLKAGRPGTSPLASKVGAPAAAPSVKPOQOOEPLAARPLG	4860
QY	4861	DLSTKDLGDPSPMDKAAXIQAQFKGYKVRKEMKQOEGPMFSHTFGDTEAQVGDALRLECV	4920
Db	4861	DLSTKDLGDPSPMDKAAXIQAQFKGYKVRKEMKQOEGPMFSHTFGDTEAQVGDALRLECV	4920
QY	4921	VASKADVBARWLKQGVETLDGRHHHIDQLGDTGTSLLIAGLDRADACGYTCQVSNKFGQV	4980
Db	4921	VASKADVBARWLKQGVETLDGRHHHIDQLGDTGTSLLIAGLDRADACGYTCQVSNKFGQV	4980
QY	4981	THSACVVVSGSSEAEASSSGGELDDAFRAARALHRLFRITKSPAESVDSBELFLSABEGPA	5040
Db	4981	THSACVVVSGSSEAEASSSGGELDDAFRAARALHRLFRITKSPAESVDSBELFLSABEGPA	5040
QY	5041	EPPEADWQYRDEDEHFCIRPEALTEARQAVTRFOEMFATLIGIVEIKLVEGQPRVEM	5100
Db	5041	EPPEADWQYRDEDEHFCIRPEALTEARQAVTRFOEMFATLIGIVEIKLVEGQPRVEM	5100
QY	5101	CISKETAPVVPPEPPLSLTSDAAPVFLTELQNVQDQGVPSFDCVVTGQMPSPVRWF	5160
Db	5101	CISKETAPVVPPEPPLSLTSDAAPVFLTELQNVQDQGVPSFDCVVTGQMPSPVRWF	5160
QY	5161	KDKKLEEDDHVMINEDQGGHOLIITAVVPADMVYRCLAENSMGVSSTKAECLRDLNLS	5220
Db	5161	KDKKLEEDDHVMINEDQGGHOLIITAVVPADMVYRCLAENSMGVSSTKAECLRDLNLS	5220
QY	5221	TDYDTAADATESSESYFSAQYLSRREQEGTESTTDEGQLPQVVEELDLQVAPGTRLAKF	5280
Db	5221	TDYDTAADATESSESYFSAQYLSRREQEGTESTTDEGQLPQVVEELDLQVAPGTRLAKF	5280
QY	5281	OLKVGYAPARLYWFKDGOPLTASAHIRMTGKILHTLBIISVTREDSQYAAIYNAMG	5340
Db	5281	OLKVGYAPARLYWFKDGOPLTASAHIRMTGKILHTLBIISVTREDSQYAAIYNAMG	5340
QY	5341	AAYSARLLVRGPEDEPEEKPAADVHEQLVPPRMLERFTPKVKYKGSISITFSVKVEGRVPP	5400
Db	5341	AAYSARLLVRGPEDEPEEKPAADVHEQLVPPRMLERFTPKVKYKGSISITFSVKVEGRVPP	5400
QY	5401	TVHMLEEABRGVJWLTGPDTPGYTVASSAQCHSLVLLDVGRQHQGYTTCIASNAAGCALC	5460
Db	5401	TVHMLEEABRGVJWLTGPDTPGYTVASSAQCHSLVLLDVGRQHQGYTTCIASNAAGCALC	5460
QY	5461	SASLHVSGULPKVBEQKVKKEALISTFLOGTTOAISAQGLETFADLGGORKEEPLAAKE	5520
Db	5461	SASLHVSGULPKVBEQKVKKEALISTFLOGTTOAISAQGLETFADLGGORKEEPLAAKE	5520
QY	5521	ALGHLSLAEVGTBEFTLQKLTQITEMVSAKIIOAKLQVPGGSDSDSKTPSPASPRHGRSR	5580
Db	5521	ALGHLSLAEVGTBEFTLQKLTQITEMVSAKIIOAKLQVPGGSDSDSKTPSPASPRHGRSR	5580
QY	5581	PSSSIOESSSESDGARGEIDFYVVTADYLPGRHQDAITLREGQYVEVLDAAHPLRW	5640
Db	5581	PSSSIOESSSESDGARGEIDFYVVTADYLPGRHQDAITLREGQYVEVLDAAHPLRW	5640
QY	5641	LVRTKTKSPSRQGVWSPAYLDRRLKLSPEWGAAPFPPEAVESEDEYKARLSVIOE	5700
Db	5641	LVRTKTKSPSRQGVWSPAYLDRRLKLSPEWGAAPFPPEAVESEDEYKARLSVIOE	5700
QY	5701	LLSSEQAFVEELOFLQSHLQHLERCPHVPIAVAGOKAVIFRNRDIGHFHSFLQELQ	5760
Db	5701	LLSSEQAFVEELOFLQSHLQHLERCPHVPIAVAGOKAVIFRNRDIGHFHSFLQELQ	5760
QY	5761	CDTDDDVAMCFIKNQAAFEQYLEFLVGRVOAESVVVSTAIQBFYKKYABEALLIAGDPSQ	5820
Db	5761	CDTDDDVAMCFIKNQAAFEQYLEFLVGRVOAESVVVSTAIQBFYKKYABEALLIAGDPSQ	5820
QY	5821	PPPLQHYLEQPVVERVQRYQALLKELIRKARNRQNCALLEQAYAVVSALPQAEANLHV	5880
Db	5821	PPPLQHYLEQPVVERVQRYQALLKELIRKARNRQNCALLEQAYAVVSALPQAEANLHV	5880
QY	5881	SLMENYPGTLEALGEBPIROGHFIVWEGAPCARPKGNHRHVFLENHVIKCPREDSRT	5940
Db	5881	SLMENYPGTLEALGEBPIROGHFIVWEGAPCARPKGNHRHVFLENHVIKCPREDSRT	5940
QY	5941	DTVSYVFRNMKLLSIDLNDQVEGDDRAFEVWQEREDSVRKYLQARTAIKSSWVKEIC	6000
Db	5941	DTVSYVFRNMKLLSIDLNDQVEGDDRAFEVWQEREDSVRKYLQARTAIKSSWVKEIC	6000
QY	6001	GIQORLALPWRPDPPEEBELADCTABGETVKLACBVTGTPKPVTSWYKDGKAVQVDPHH	6060
Db	6001	GIQORLALPWRPDPPEEBELADCTABGETVKLACBVTGTPKPVTSWYKDGKAVQVDPHH	6060
QY	6061	ILIEDPDGSCALITLDTGVDSGQVCFAAASAGNCSTLGLKILVQVPRPFVNVKVRASPFV	6120

```

Db 6061 ILIEDPGSCALILDSITGVDSGQYMCFAASAAGNCSTLGKILVQPPRVNKLSPFV 6120
Qy 6121 EGEDAQFTCTTEGAPYQIRWYKDGALLTGNKQPTLSEPRSGLLVLVIRAAKEDLGLY 6180
Db 6121 EGEDAQFTCTTEGAPYQIRWYKDGALLTGNKQPTLSEPRSGLLVLVIRAAKEDLGLY 6180
Qy 6181 BCELVNRLGSSARASAEIRIQSPMLQAOEQCHREQLVAIVEDTTLER-----ADQEV 6232
Db 6181 BCELVNRLGSSARASAEIRIQSPMLQAOEQCHREQLVAIVEDTQETKVPKTVIIETIT 6240
Qy 6233 SVLKRLGPKAPGSTGDLTGPGCPRG-----APAL-----OETSQP- 6271
Db 6241 TVKSPRQRR-SPSKSPRSRCSASPLRPGLLAPOLLVLPAGAGPRRPEAPGQKPV 6299
Qy 6272 -PVTGTSEAPVPRV-----PQPLLHEGPE-----QPEAIARAQEWTPIRMEGA 6317
Db 6300 VPTLYVTEAEASHSPALPGSLGPGPKWVEVEETIEVRVKKMGPGQVSPTE--VPRSSSGH 6357
Qy 6318 AW--PGACTG 6325
Db 6358 LFTLPGATPG 6367

RESULT 4
US-10-077-130-2
; Sequence 2, Application US/10077130
; Publication No. US20020169742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Thereof
; FILE REFERENCE: MPI2001-047P1RCP1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2630
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-077-130-2

Query Match 33.2%; Score 13710; DB 13; Length 2630;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2630; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5339 MGAAYSSARLLVRCGPDEPEEKPSADVHEQLVPPRMLERFTPKVKYKSSITFSVKVEGRP 5398
Db 1 MGAAYSSARLLVRCGPDEPEEKPSADVHEQLVPPRMLERFTPKVKYKSSITFSVKVEGRP 60

Qy 5399 VPTVHMLREEAERGLVIGDPTPGYTVASSAQHSLVLLDVGRHQHGYTTCIASNAAGQA 5458
Db 61 VPTVHMLREEAERGLVIGDPTPGYTVASSAQHSLVLLDVGRHQHGYTTCIASNAAGQA 120

Qy 5459 LCSASLVHSGLPKVEEKEKVEALISTFLOGTTOAISAOGLETFASADLGGQKEEPLAA 5518
Db 121 LCSASLVHSGLPKVEEKEKVEALISTFLOGTTOAISAOGLETFASADLGGQKEEPLAA 180

Qy 5519 KEALCHLSLAVGTEEFLOKLTSTOITWVSAKITQAKLQVPGGSDSDSKTPSASPRHGR 5578
Db 181 KEALCHLSLAVGTEEFLOKLTSTOITWVSAKITQAKLQVPGGSDSDSKTPSASPRHGR 240

Qy 5579 SRPSSSQESSSESDGDARGEIYVVTADYLPFGAEQDAILREGQVVEVLDAAHPL 5638
Db 241 SRPSSSQESSSESDGDARGEIYVVTADYLPFGAEQDAILREGQVVEVLDAAHPL 300

Qy 5639 RNLVTRKPTKSSPRQGVSPAYLDRRLKLSPEWGAAEAPEFFPGEAVSEDEYKARLSSVI 5698
Db 301 RNLVTRKPTKSSPRQGVSPAYLDRRLKLSPEWGAAEAPEFFPGEAVSEDEYKARLSSVI 360

```

```

Qy 5699 QELLSEQAFVEELOFLOSHHLQHLERCPHPVPIAVAGQKAVIFRNVRDIGRHFHSFLQEL 5758
Db 361 QELLSEQAFVEELOFLOSHHLQHLERCPHPVPIAVAGQKAVIFRNVRDIGRHFHSFLQEL 420
Qy 5759 QQCDDDDDDVAMCFIKNOAAFEQYLEFLVGRVOABSVVSTAIQBFYKKYABEALLAGDPS 5818
Db 421 QQCDDDDDDVAMCFIKNOAAFEQYLEFLVGRVOABSVVSTAIQBFYKKYABEALLAGDPS 480
Qy 5819 QPPPPLOHYLEOPVERVORYOALLKELIRKANRQNCALLEQAYAVVSALPORAENKL 5878
Db 481 QPPPPLOHYLEOPVERVORYOALLKELIRKANRQNCALLEQAYAVVSALPORAENKL 540
Qy 5879 HVSLMENYPGTLEALGEPIRQGHFIWEGAPGASMPKGNHRHVFLEFNHLVICRPRDS 5938
Db 541 HVSLMENYPGTLEALGEPIRQGHFIWEGAPGASMPKGNHRHVFLEFNHLVICRPRDS 600
Qy 5939 RDTVSVYFRNMKLSISDINDQVEGDDRAFEVWQEREDSVRKYLLQARTAIKSSWYKE 5998
Db 601 RDTVSVYFRNMKLSISDINDQVEGDDRAFEVWQEREDSVRKYLLQARTAIKSSWYKE 660
Qy 5999 ICGIQORLALPWRPDPPEEBELADCTAELGETVKLACRVGTGTPKPVISWYKDGKAVQVDP 6058
Db 661 ICGIQORLALPWRPDPPEEBELADCTAELGETVKLACRVGTGTPKPVISWYKDGKAVQVDP 720
Qy 6059 KHILIEDPGSCALILDSITGVDSGQYMCFAASAAGNCSTLGKILVQPPRVNKLSPFV 6118
Db 721 KHILIEDPGSCALILDSITGVDSGQYMCFAASAAGNCSTLGKILVQPPRVNKLSPFV 780
Qy 6119 FVEGEDAQFTCTTEGAPYQIRWYKDGALLTGNKQPTLSEPRSGLLVLVIRAAKEDLGL 6178
Db 781 FVEGEDAQFTCTTEGAPYQIRWYKDGALLTGNKQPTLSEPRSGLLVLVIRAAKEDLGL 840
Qy 6179 LYECELVNRLGSSARASAEIRIQSPMLQAOEQCHREQLVAAVEDTTLERADQEVTSVLKRL 6238
Db 841 LYECELVNRLGSSARASAEIRIQSPMLQAOEQCHREQLVAAVEDTTLERADQEVTSVLKRL 900
Qy 6239 LGPKAPGSTGDLTGPGCPRGAPALQETGSGPPVTCTSEAPVPPVPQPLLHEGPEQE 6298
Db 901 LGPKAPGSTGDLTGPGCPRGAPALQETGSGPPVTCTSEAPVPPVPQPLLHEGPEQE 960
Qy 6299 PEAIARAQEWTPVIRMEGAAMPAGCTGELLMDVHSHVVRETTQRTTYQAIDTHTARPPS 6358
Db 961 PEAIARAQEWTPVIRMEGAAMPAGCTGELLMDVHSHVVRETTQRTTYQAIDTHTARPPS 1020
Qy 6359 MOVTTIEDVQAQGTGTAQFEALIEGDPQPSVTWYKDSVQLVDSTRLSQOQEGTTYSVLVRH 6418
Db 1021 MOVTTIEDVQAQGTGTAQFEALIEGDPQPSVTWYKDSVQLVDSTRLSQOQEGTTYSVLVRH 1080
Qy 6419 VASKDAGVYVTCIAQNTGGQVLCKAELLVLGGDNEPDSKQSHRRKLHSFYEVKEEIGRGV 6478
Db 1081 VASKDAGVYVTCIAQNTGGQVLCKAELLVLGGDNEPDSKQSHRRKLHSFYEVKEEIGRGV 1140
Qy 6479 FGFVKRVQHKNGKILCAAKFIPLRSRTRAQYRREDILAAHSHPLVTGLLDQFTRKTLI 6538
Db 1141 FGFVKRVQHKNGKILCAAKFIPLRSRTRAQYRREDILAAHSHPLVTGLLDQFTRKTLI 1200
Qy 6539 LLELCSSPELLDLRYRKGVTAEVKYIQQLEGLHYLHSHGVHLHDIKPSNLMVHP 6598
Db 1201 LLELCSSPELLDLRYRKGVTAEVKYIQQLEGLHYLHSHGVHLHDIKPSNLMVHP 1260
Qy 6599 AREDIKICDFGAQNTIPAEILQFSGYSGPFSPEIQQONPVSEASDIWANGVTSYLSLT 6658
Db 1261 AREDIKICDFGAQNTIPAEILQFSGYSGPFSPEIQQONPVSEASDIWANGVTSYLSLT 1320
Qy 6659 CSSPAGESDRTATLVLEGRVSWSSPMAHLSDAXDFIKATLQAPQAPPSAACQLSH 6718
Db 1321 CSSPAGESDRTATLVLEGRVSWSSPMAHLSDAXDFIKATLQAPQAPPSAACQLSH 1380
Qy 6719 PWFLLKMPAEAEHAFINTKQLKFLLRARSWORSLMSYKSIYVMSIPELLRGPPDPSLGV 6778
Db 1381 PWFLLKMPAEAEHAFINTKQLKFLLRARSWORSLMSYKSIYVMSIPELLRGPPDPSLGV 1440

```

QY 6779 ABHLCDTGGSSSSSSDNELAPARAKSLPPSPVTHSPHLLHPRGFLRPSASIPERABEA 6938  
DB 1441 ABHLCDTGGSSSSSSDNELAPARAKSLPPSPVTHSPHLLHPRGFLRPSASIPERABEA 1500  
QY 6839 SERSTEAPAPASPGAGAPPAAGCVPRHVSIRSLFYHQAGSEPHGALAPGSRHHPARR 6898  
DB 1501 SERSTEAPAPASPGAGAPPAAGCVPRHVSIRSLFYHQAGSEPHGALAPGSRHHPARR 1560  
QY 6899 RHLKGGYIAGALPGLRPLMEHRVLEBEAREQATLLAKAPSFETALRLPAPSGTHLAP 6958  
DB 1561 RHLKGGYIAGALPGLRPLMEHRVLEBEAREQATLLAKAPSFETALRLPAPSGTHLAP 1620  
QY 6959 GHSHLEHDSSTPRPSEACGEAQLPSAPSGAPIRDMGHPOGSKQLPSTGHPGTQAQ 7018  
DB 1621 GHSHLEHDSSTPRPSEACGEAQLPSAPSGAPIRDMGHPOGSKQLPSTGHPGTQAQ 1680  
QY 7019 PERSPDPSWQOPAPFCHPKQGSAPQECSPHPAVAPCPGSPFPFGCKEAPLVPSFFL 7078  
DB 1681 PERSPDPSWQOPAPFCHPKQGSAPQECSPHPAVAPCPGSPFPFGCKEAPLVPSFFL 1740  
QY 7079 GQOPAPPAPAKASPLDSKMGPGDISLGRPKGPCSPGSGASQASSQVSSLRVGSSQV 7138  
DB 1741 GQOPAPPAPAKASPLDSKMGPGDISLGRPKGPCSPGSGASQASSQVSSLRVGSSQV 1800  
QY 7139 GTEPGPSLDAEGWTOAEEDLSGSTPTLQRPQEQATMRKFSLGGRGYAGVAGYGTFAFGG 7198  
DB 1801 GTEPGPSLDAEGWTOAEEDLSGSTPTLQRPQEQATMRKFSLGGRGYAGVAGYGTFAFGG 1860  
QY 7199 DAGMLOGGPMWARIANAVSSEBEQBEAREAQSEBQOQARAESELPQVSARPPVEVG 7258  
DB 1861 DAGMLOGGPMWARIANAVSSEBEQBEAREAQSEBQOQARAESELPQVSARPPVEVG 1920  
QY 7259 RAPTRSSPEPTWEDIGQVSLVQIRDLSDGDAEADTISLDISEVDPAVLNLSLDIYKYL 7318  
DB 1921 RAPTRSSPEPTWEDIGQVSLVQIRDLSDGDAEADTISLDISEVDPAVLNLSLDIYKYL 1980  
QY 7319 PFEMIFPKVXKSAQPEPPSPWABEEAEAEFFPEPTWMPGELGPHAGLEITEESESDVDALL 7378  
DB 1981 PFEMIFPKVXKSAQPEPPSPWABEEAEAEFFPEPTWMPGELGPHAGLEITEESESDVDALL 2040  
QY 7379 AEAAGRKRWSSPSRSIFHPGPHLPDDEPAEJGLERVKASVEHISRLIKGRPEGLEK 7438  
DB 2041 AEAAGRKRWSSPSRSIFHPGPHLPDDEPAEJGLERVKASVEHISRLIKGRPEGLEK 2100  
QY 7439 EGPRKPKGLASFRISGLKSDRAPTFLRELSDETIVLGQSVTLACQVSAQPAQAATWSK 7498  
DB 2101 EGPRKPKGLASFRISGLKSDRAPTFLRELSDETIVLGQSVTLACQVSAQPAQAATWSK 2160  
QY 7499 DGAPLESSRVLISATLKNFOLLITLVVAEDLGYYTCSVSNALGTVTTCVILKARPS 7558  
DB 2161 DGAPLESSRVLISATLKNFOLLITLVVAEDLGYYTCSVSNALGTVTTCVILKARPS 2220  
QY 7559 SSPCPDICEVADGVLLVWKPVESYGPVTYIVQCSLEGSMWTLASDIFDCCYLTSKLSR 7618  
DB 2221 SSPCPDICEVADGVLLVWKPVESYGPVTYIVQCSLEGSMWTLASDIFDCCYLTSKLSR 2280  
QY 7619 GGTYTFRACVSKAGMGYSPSPSOVLILGGPSHLASEEESQGRSAQPLPSTKTAFOTQI 7678  
DB 2281 GGTYTFRACVSKAGMGYSPSPSOVLILGGPSHLASEEESQGRSAQPLPSTKTAFOTQI 2340  
QY 7679 QRGRESVVRQWEKASGRALAAKIIPYHPKDKTAVLREYEAALKGLRPHLAQLHAAYLSP 7738  
DB 2341 QRGRESVVRQWEKASGRALAAKIIPYHPKDKTAVLREYEAALKGLRPHLAQLHAAYLSP 2400  
QY 7739 RHLVLILELCSGPELLPCLAPRASYSSEVKDYLMQWLSATQYLNHQLHLDLRSENMI 7798  
DB 2401 RHLVLILELCSGPELLPCLAPRASYSSEVKDYLMQWLSATQYLNHQLHLDLRSENMI 2460  
QY 7799 ITEYNLLKVDLGNQAQSLQSKVLPSPDKFKDYLETMAPELLEGQVPTDIIWAGYTA 7858  
DB 2461 ITEYNLLKVDLGNQAQSLQSKVLPSPDKFKDYLETMAPELLEGQVPTDIIWAGYTA 2520  
QY 7859 IMLSAEYPVSEGGADLQRLKGLVRLSRCYAGLSGGAVAFRLSTLCAQPGWRPCASSC 7918

DB 2521 IMLSAEYPVSEGGADLQRLKGLVRLSRCYAGLSGGAVAFRLSTLCAQPGWRPCASSC 2580  
QY 7919 LQCPWLTERGPGACSPAPVTFPTARLVRVFRNRERKRALLYKRHNLAQVR 7968  
DB 2581 LQCPWLTERGPGACSPAPVTFPTARLVRVFRNRERKRALLYKRHNLAQVR 2630  
RESULT 5  
US-10-307-019-6  
; Sequence 6, Application US/10307019  
; Publication No. US20030108533A1  
; GENERAL INFORMATION:  
; APPLICANT: Zeng, Wenlin  
; APPLICANT: Stanton, Lawrence  
; APPLICANT: SCIOS, INC.  
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION  
; FILE REFERENCE: SCIOS.021DV1  
; CURRENT APPLICATION NUMBER: US/10/307,019  
; CURRENT FILING DATE: 2002-11-26  
; PRIOR APPLICATION NUMBER: 09/548,473  
; PRIOR FILING DATE: 2000-04-13  
; PRIOR APPLICATION NUMBER: 60/129,552  
; PRIOR FILING DATE: 1999-04-16  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 2596  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-307-019-6  
Query Match 32.8%; Score 13528; DB 14; Length 2596;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 5373 MLERFPPKKYKKGSSITFSVKVEGRVPVTHMLREBAERGVLMWIGDPDTPGYTVASSAQOH 5432  
DB 1 MLERFPPKKYKKGSSITFSVKVEGRVPVTHMLREBAERGVLMWIGDPDTPGYTVASSAQOH 60  
QY 5433 SLVLLDVGHQHGQYTCIASNAAGQALCSASLHVSGLPKVEQEKVKEALISTFLOGTTQ 5492  
DB 61 SLVLLDVGHQHGQYTCIASNAAGQALCSASLHVSGLPKVEQEKVKEALISTFLOGTTQ 120  
QY 5493 AISAQOLETASFADLGGQKKEPPLAAKEALGHLSLAEVGTEBFLOKLTQSITEMVSAKIT 5552  
DB 121 AISAQOLETASFADLGGQKKEPPLAAKEALGHLSLAEVGTEBFLOKLTQSITEMVSAKIT 180  
QY 5553 QAKLVPGGSDSDSKTPSASPRHGRSRSPSSIOESSSESDGDARCEIFDIYVVVTADYL 5612  
DB 181 QAKLVPGGSDSDSKTPSASPRHGRSRSPSSIOESSSESDGDARCEIFDIYVVVTADYL 240  
QY 5613 PLGAEQDAITLREGQVVEVLDAAPLRLVRLTKPTKSSPSRQGWSPAYLDRLKLSPW 5672  
DB 241 PLGAEQDAITLREGQVVEVLDAAPLRLVRLTKPTKSSPSRQGWSPAYLDRLKLSPW 300  
QY 5673 GAABEPFPGAEVSEDEYKARLSSVTQELLSSEQAQVBELOQLQSHHLQHLRCHPVTIA 5732  
DB 301 GAABEPFPGAEVSEDEYKARLSSVTQELLSSEQAQVBELOQLQSHHLQHLRCHPVTIA 360  
QY 5733 VAGQKAVIRNRVDIRGFHSSFIQELQQCDTDDDVAMCFIKQAQAFQVLEFLVGRVQAE 5792  
DB 361 VAGQKAVIRNRVDIRGFHSSFIQELQQCDTDDDVAMCFIKQAQAFQVLEFLVGRVQAE 420  
QY 5793 SVVVTATIQEFYKKAABEALLAGDPSPQPPPLQHYLEQVFRVORYQALLKELIRNKAR 5852  
DB 421 SVVVTATIQEFYKKAABEALLAGDPSPQPPPLQHYLEQVFRVORYQALLKELIRNKAR 480  
QY 5853 NRQNCALLQCAVAVSALPQRAENKLHVSIMENYPTLEALGEPTRQGHFIWEGAPGAR 5912  
DB 481 NRQNCALLQCAVAVSALPQRAENKLHVSIMENYPTLEALGEPTRQGHFIWEGAPGAR 540  
QY 5913 MPWKGNHRVFLFRNHLVICKPRRSRTDTVSVFRNMMLSSIDLNDQVEGDDRAFEVW 5972

Db 541 MPWKGHNRHVLFRNHLVCKPRDRSDTDSVYFRNMKLSLIDLNQVGGDRAFVW 600  
 Qy 5973 QEREDSVRKYLQARTAIKSSWKEICGQOORLALPVRPPDPFEELADCTAELGETVK 6032  
 Db 601 QEREDSVRKYLQARTAIKSSWKEICGQOORLALPVRPPDPFEELADCTAELGETVK 660  
 Qy 6033 LACVTCGTPKVIISWYKDGKAVQVDPHILLIEDPDGSCALILDSLTGSDSQWMCFAASA 6092  
 Db 661 LACVTCGTPKVIISWYKDGKAVQVDPHILLIEDPDGSCALILDSLTGSDSQWMCFAASA 720  
 Qy 6093 AGNSTGLKILVPPFRFVNVKVRASPFVEGEDAQTCTIEGAPYQIRWYKDGALLTGN 6152  
 Db 721 AGNSTGLKILVPPFRFVNVKVRASPFVEGEDAQTCTIEGAPYQIRWYKDGALLTGN 780  
 Qy 6153 KFOTLSEPRSGLLVIRAAKEDLGIYECELVNRIGSARASAEALRQSPMLQAOQCHR 6212  
 Db 781 KFOTLSEPRSGLLVIRAAKEDLGIYECELVNRIGSARASAEALRQSPMLQAOQCHR 840  
 Qy 6213 EQLVAIVEDTTLERADOEVTSVLRLLGPKAPGSTGDLTGPGRGAPALQETGSQPP 6272  
 Db 841 EQLVAIVEDTTLERADOEVTSVLRLLGPKAPGSTGDLTGPGRGAPALQETGSQPP 900  
 Qy 6273 VTGTSEAPVPRVPOQLLHEGPEQBEAARAQEWTPVIRMEGAAMPAGAGTGELLWDVH 6332  
 Db 901 VTGTSEAPVPRVPOQLLHEGPEQBEAARAQEWTPVIRMEGAAMPAGAGTGELLWDVH 960  
 Qy 6333 SHVRETTQRTYTYCAIDTHTARPPSMQVITIEDVQAGTGAQPEALIEGDPQPSVTWYK 6392  
 Db 961 SHVRETTQRTYTYCAIDTHTARPPSMQVITIEDVQAGTGAQPEALIEGDPQPSVTWYK 1020  
 Qy 6393 DSQVLVSTRLSQOQEGTYSVLIRHVASKDAGVYTCIAQNTGGQVLCKAELIIVLGDNE 6452  
 Db 1021 DSQVLVSTRLSQOQEGTYSVLIRHVASKDAGVYTCIAQNTGGQVLCKAELIIVLGDNE 1080  
 Qy 6453 PDSEKQSHRRKLSHFYEVKEEIGRGVFKVQVHKGKILCAAKTIPLESRTAQAIRE 6512  
 Db 1081 PDSEKQSHRRKLSHFYEVKEEIGRGVFKVQVHKGKILCAAKTIPLESRTAQAIRE 1140  
 Qy 6513 RDILAALSHPLVTLGLDQFETKTLILILELCSSEELLDRLYRGVVTEAEVYIQQLV 6572  
 Db 1141 RDILAALSHPLVTLGLDQFETKTLILILELCSSEELLDRLYRGVVTEAEVYIQQLV 1200  
 Qy 6573 EGLYLSHSHVGLHLDIKPSNLMVHPAREDIKICDFGAQNIPTAELOFQSYGSPFVSP 6632  
 Db 1201 EGLYLSHSHVGLHLDIKPSNLMVHPAREDIKICDFGAQNIPTAELOFQSYGSPFVSP 1260  
 Qy 6633 EIIQONPVSEASDIWANGVISYLSLTCSPPAGSDRATLNLVLEGRVSSPMAAHLSE 6692  
 Db 1261 EIIQONPVSEASDIWANGVISYLSLTCSPPAGSDRATLNLVLEGRVSSPMAAHLSE 1320  
 Qy 6693 DAKOFIKATLQAPQAPSAQAQCLSHFWFLKSPABEAHFINTKQLFLLARSRWORSIM 6752  
 Db 1321 DAKOFIKATLQAPQAPSAQAQCLSHFWFLKSPABEAHFINTKQLFLLARSRWORSIM 1380  
 Qy 6753 SYKSILVMSRIPPELLRPPDPSLGVARHLCDRTGGSSSSSSSDNELAPFAAKSLPPS 6812  
 Db 1381 SYKSILVMSRIPPELLRPPDPSLGVARHLCDRTGGSSSSSSSDNELAPFAAKSLPPS 1440  
 Qy 6813 PVTHSPLHPRGFLRPSASLPEEAERSTEAAPPASPEGAGPPAAQOCVPRHSVIRS 6872  
 Db 1441 PVTHSPLHPRGFLRPSASLPEEAERSTEAAPPASPEGAGPPAAQOCVPRHSVIRS 1500  
 Qy 6873 LFYHQGESPEHGALAPCSRHPARRRLLKGGYIAGALPGLREPLMEHRVLEEAAREE 6932  
 Db 1501 LFYHQGESPEHGALAPCSRHPARRRLLKGGYIAGALPGLREPLMEHRVLEEAAREE 1560  
 Qy 6933 QATLLAKAPSETALRUPASGTHLAPGSHSLEHDSPTPRPSEACGEAQRLPSPSGG 6992  
 Db 1561 QATLLAKAPSETALRUPASGTHLAPGSHSLEHDSPTPRPSEACGEAQRLPSPSGG 1620  
 Qy 6993 APIRDMGHPOGSKQLPSTGGHPTGAQPERSPDPSWGPQAPFCHPKQGSAPQSGCSHPA 7052

Db 1621 APIRDMGHPOGSKQLPSTGGHPTGAQPERSPDPSWGPQAPFCHPKQGSAPQSGCSHPA 1680  
 Qy 7053 VAPCPGSGFPFGSCKEARLVPSSPFLGQOPAPPAPAKASPLDLSKMGPGDLSLGRPKPG 7112  
 Db 1681 VAPCPGSGFPFGSCKEARLVPSSPFLGQOPAPPAPAKASPLDLSKMGPGDLSLGRPKPG 1740  
 Qy 7113 PCSSPGSAQASSSSQVSLRVGSSQVTEPGPSLDAEGWTOEAEDLSSTFTLQRPQEA 7172  
 Db 1741 PCSSPGSAQASSSSQVSLRVGSSQVTEPGPSLDAEGWTOEAEDLSSTFTLQRPQEA 1800  
 Qy 7173 TMRKPSLGRGGYAGVAGYGTAFGGDAGMGLGQPMWARIAWAVSSEEEQEEAARAS 7232  
 Db 1801 TMRKPSLGRGGYAGVAGYGTAFGGDAGMGLGQPMWARIAWAVSSEEEQEEAARAS 1860  
 Qy 7233 QSEEEQEEAARASPLPOVSARPVPEVGRAPTRSSPEPTWEDIGOVSLVQIRDLSDGAEEA 7292  
 Db 1861 QSEEEQEEAARASPLPOVSARPVPEVGRAPTRSSPEPTWEDIGOVSLVQIRDLSDGAEEA 1920  
 Qy 7293 DTISLDSIENVDPAYLNLSDLYDIKYLPPEFMI FRKVPKSAQPEPPSPMAESELAEFEPT 7352  
 Db 1921 DTISLDSIENVDPAYLNLSDLYDIKYLPPEFMI FRKVPKSAQPEPPSPMAESELAEFEPT 1980  
 Qy 7353 WPMGELGPHAGLEITTESEEDVDALLABAAVGRKRKSSPSRSLFHPGRHLPLDEPAEL 7412  
 Db 1981 WPMGELGPHAGLEITTESEEDVDALLABAAVGRKRKSSPSRSLFHPGRHLPLDEPAEL 2040  
 Qy 7413 GLRERVKASVEHISRLKGRPEGLEKEGPPRKPGLASFRLSGLKSWDRAPFTFLRELSDE 7472  
 Db 2041 GLRERVKASVEHISRLKGRPEGLEKEGPPRKPGLASFRLSGLKSWDRAPFTFLRELSDE 2100  
 Qy 7473 TVVLQOSTVLACQVSAQAAQATWSKCAPLESSRVLISATLKNFOLLTLVWVAEDIG 7532  
 Db 2101 TVVLQOSTVLACQVSAQAAQATWSKCAPLESSRVLISATLKNFOLLTLVWVAEDIG 2160  
 Qy 7533 VYTCVSVALGTVTTTGVLRKAERPSSPCPDIGEVADGVLVWKPVESYGPVTYIVQC 7592  
 Db 2161 VYTCVSVALGTVTTTGVLRKAERPSSPCPDIGEVADGVLVWKPVESYGPVTYIVQC 2220  
 Qy 7593 SLEGGSWTTLASDIPDCCYLTSKLSRGTYTFRACVSKAGMGYPSPSEQVLLGGPSHL 7652  
 Db 2221 SLEGGSWTTLASDIPDCCYLTSKLSRGTYTFRACVSKAGMGYPSPSEQVLLGGPSHL 2280  
 Qy 7653 ASEESQGRSQAPLSTTKTFAFQTOIGRGRSVVRQCKEASGRALAAKIIPYHPKDKTA 7712  
 Db 2281 ASEESQGRSQAPLSTTKTFAFQTOIGRGRSVVRQCKEASGRALAAKIIPYHPKDKTA 2340  
 Qy 7713 VLREYALKGLRHPHQAHLAAYLSPRHLVILILELCSGPELLPCLAEASYSSEVSKDYL 7772  
 Db 2341 VLREYALKGLRHPHQAHLAAYLSPRHLVILILELCSGPELLPCLAEASYSSEVSKDYL 2400  
 Qy 7773 WMLSATQYLHNOHTLHLDLRSENMIITEYNLLKVVLDLGNACQSLQSKVLPDSKFKDYLE 7832  
 Db 2401 WMLSATQYLHNOHTLHLDLRSENMIITEYNLLKVVLDLGNACQSLQSKVLPDSKFKDYLE 2460  
 Qy 7833 TMAPELLSSQGAVPOTDIWAGTAFIMLSAEYVPSSESGARDLQGRKGLVRLSRVRE 7892  
 Db 2461 TMAPELLSSQGAVPOTDIWAGTAFIMLSAEYVPSSESGARDLQGRKGLVRLSRVRE 2520  
 Qy 7893 LSGGAVAFRLSTLCAQWGRCCASSCLOCPLWTEGGRACSPAPVPTPTABLRVVRRE 7952  
 Db 2521 LSGGAVAFRLSTLCAQWGRCCASSCLOCPLWTEGGRACSPAPVPTPTABLRVVRRE 2580  
 Qy 7953 KERALLYKXRNHLAQR 7968  
 Db 2581 KERALLYKXRNHLAQR 2596

RESULT 6

US-10-093-463-72

; Sequence 72, Application US/10093463

; Publication No. US20030208039A1

; GENERAL INFORMATION:

; APPLICANT: Padigaru, Muralidhara

```
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glenda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zernusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malyankar, Uriel
; APPLICANT: Anderson, David
; APPLICANT: Patturajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; TITLE OF INVENTION: No. US20030208039A1el Antibodies that Bind to Antigenic Polypeptide
; FILE REFERENCE: 21402-290A (Cura 530AT)
; CURRENT APPLICATION NUMBER: US/10/093,463
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675
; PRIOR FILING DATE: 2001-04-14
; PRIOR APPLICATION NUMBER: 60/338,092
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 72

; LENGTH: 4691
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-093-463-72

Query Match      25.5%; Score 10519.5; DB 15; Length 4691;
Best Local Similarity 51.1%; Pred. No. 0;
Matches 2367; Conservative 118; Mismatches 342; Indels 1807; Gaps 66;

QY      2595  MPLYNDSPHEISHKGRHRTLVLKSIQRADAGIVRASSLKVSTASARLEVRVKVFWFLKALD 2654
Db      1  MPLYNDSPHEISHKGRHRTLVLKSIQRADAGIVRASSLKVSTASARLEVRVKVFWFLKALD 60

QY      2655  DLSAEERGTLALQCEVSDPEAHVVRKQGVQGPSKDYDLHTAGTRGLVVDVSPEDAG 2714
Db      61  DLSAEERGTLALQCEVSDPEAHVVRKQGVQGPSKDYDLHTAGTRGLVVDVSPEDAG 120

QY      2715  LYTCHVGSSETRARVRVHDLHVGITKRLKTMEVLGEGSCSFECVLSHESASDPAMWTVG 2774
Db      121  LYTCHVGSSETRARVRVHDLHVGITKRLKTMEVLGEGSCSFECVLSHESASDPAMWTVG 180

QY      2775  KTVGSSSRFOATRQGRKYILVVRKAAPS DAGEWVFSVRLGTSKASLIVREERPAAIKPLE 2834
Db      181  KTVGSSSRFOATRQGRKYILVVRKAAPS DAGEWVFSVRLGTSKASLIVREERPAAIKPLE 240

QY      2835  DQWVAFGEDVVELRCELSRAGTPVHMLKORKATRKQKDYDVCCEGTAMMLVIRGASLKDAG 2894
Db      241  DQWVAFGEDVVELRCELSRAGTPVHMLKORKATRKQKDYDVCCEGTAMMLVIRGASLKDAG 300

QY      2895  EYTCEVEASKSTASLHVEEKANCFTEELTNLOVEKGTAVFTCKTEHPAATVTRKGLLE 2954
Db      301  EYTCEVEASKSTASLHVEEKANCFTEELTNLOVEKGTAVFTCKTEHPAATVTRKGLLE 360

QY      2955  LRASGKHQPSQGLTLRLTISALEKADSDTYTCDTGOAQSRAQLLVQGRVHIIDEDLV 3014
Db      361  LRASGKHQPSQGLTLRLTISALEKADSDTYTCDTGOAQSRAQLLVQGRVHIIDEDLV 420

QY      3015  DVQEGSSATFRCRISPANYEPVHFLDKTPLHANELNEIDAOPGGVHVLTLLQALAKDSG 3074
Db      421  DVQEGSSATFRCRISPANYEPVHFLDKTPLHANELNEIDAOPGGVHVLTLLQALAKDSG 480

QY      3075  TIYFEAGDORASAAALRVTEKPSVFSRELTDAITIGEDLTLCETSTCDIPMCWTKDGT 3134
Db      481  TIYFEAGDORASAAALRVTEKPSVFSRELTDAITIGEDLTLCETSTCDIPMCWTKDGT 540

QY      3135  LRGSARCOLSHGHRRAQLLITGATLQDSGRYKCEAGGACSSIVRVHARVPVFOEALKDL 3194
Db      541  LRGSARCOLSHGHRRAQLLITGATLQDSGRYKCEAGGACSSIVRVHARVPVFOEALKDL 600

QY      3195  EVLEGGAATLRCVLSVAAPVKWCGNNVLRPGDKYSLRQEGAMLELVVRNLRPODSGRY 3254
Db      601  EVLEGGAATLRCVLSVAAPVKWCGNNVLRPGDKYSLRQEGAMLELVVRNLRPODSGRY 660

QY      3255  SCSFGDQTTSATLTVTALPAQFIGKLRNKEATGATATLRCBELSKTAPVWNRKGSSTLRD 3314
Db      661  SCSFGDQTTSATLTVTALPAQFIGKLRNKEATGATATLRCBELSKTAPVWNRKGSSTLRD 720

QY      3315  GDRYCLRQDGAMCELQIRGLAMVDAAEYSCVCGEERTSASLTIRPMPAHFICRLRHQESI 3374
Db      721  GDRYCLRQDGAMCELQIRGLAMVDAAEYSCVCGEERTSASLTIRPMPAHFICRLRHQESI 780

QY      3375  EGATATLRCBELSKAAPVWNRKGRESLRDGRHSRLQDGVAVCELQICGLAVADAGEYSCVC 3434
Db      781  EGATATLRCBELSKAAPVWNRKGRESLRDGRHSRLQDGVAVCELQICGLAVADAGEYSCVC 840

QY      3435  GEERTSATLTVRKALPAKFTTEGLRNEEAVEGATAMLWCELSKVAPVWNRKGPENLRDGRY 3494
Db      841  GEERTSATLTVRKALPAKFTTEGLRNEEAVEGATAMLWCELSKVAPVWNRKGPENLRDGRY 900

QY      3495  ILRQEGTRCELOICGLAMADAGEYLCVCCQERTSATLTIRALPARFIEDVKNQOEAREGAT 3554
Db      901  ILRQEGTRCELOICGLAMADAGEYLCVCCQERTSATLTIRALPARFIEDVKNQOEAREGAT 960
```



3555	QY	AVLQCELSAAPVEMWKGSETLRDGDYSLRDQGTKELOIRGLAMADTGEYSCVCGQER	3614
961	Db	AVLQCELSAAPVEMWKGSETLRDGDYSLRDQGTKELOIRGLAMADTGEYSCVCGQER	1020
3615	QY	TSAMLTVRALPIKFTGELRNEEATGATAVLRCELSKVAPEVWKGHETLRDGDHSLRQ	3674
1021	Db	TSAMLTVRALPIKFTGELRNEEATGATAVLRCELSKVAPEVWKGHETLRDGDHSLRQ	1080
3675	QY	DGARCELQIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPSKFI EGLRNEEATGDTATLW	3734
1081	Db	DGARCELQIRGLVAEDAGEYLCMCGKERTSAMLTVRAMPSKFI EGLRNEEATGDTATLW	1140
3735	QY	CELSKAAPVEMWKGHETLRDGDHSLRDQGSRCCELOIRGLAVVDAGEYSCVCCQERTSAT	3794
1141	Db	CELSKAAPVEMWKGHETLRDGDHSLRDQGSRCCELOIRGLAVVDAGEYSCVCCQERTSAT	1200
3795	QY	LTVRALPARFIEDVKNOEAREGATAVLQCELSKAAPVEMWKGSETLRGGDRYSLRDQDTR	3854
1201	Db	LTVRALPARFIEDVKNOEAREGATAVLQCELSKAAPVEMWKGSETLRGGDRYSLRDQDTR	1260
3855	QY	CELOIHGLSVADTGEYSCVCGQERTSATLTVR	3886
1261	Db	CELOIHGLSVADTGEYSCVCGQERTSATLTVRALPARFTQDLTKYKASEGATATLQCELS	1320
3887	QY	-----	3886
1321	Db	KVAPVEMWKGPELTRDGGYSLRKQDGTCELOIHDLISVADAGEYSCMCGQERTSATLTVR	1380
3887	QY	-----	3886
1381	Db	ALPARFTEGLRNEEAMEGATATLQCELSKAAPVEMWKGLEALRDGDKYSLRDQGVACELQ	1440
3887	QY	-----	3886
1441	Db	IHGLAMADNGYSCVCGQERTSATLTVRALPARFIEDMRNOKATEGATVTLQCKLRKAAP	1500
3887	QY	-----	3886
1501	Db	VEWRKGPNTLKDGGDRYSLRKQDGTSCCELOIRGLVIADAGEYSCICEQERTSATLTVRALPA	1560
3887	QY	-----	3886
1561	Db	RFIEDVRNHEATGATAVLQCELSKAAPVEMWKGSETLRDGDYSLRDQGTCELOIRGL	1620
3887	QY	-----	3886
1621	Db	AVEDTGEYLCVCGQERTSATLTVRALPARFIDNMTNQAREGATATLHCELSKVAPEVWR	1680
3887	QY	-----	3886
1681	Db	KGPETLRDGDHSLRDQGSRCCELOIRGLAVVDAGEYSCVCGQERTSATLTVRALPARFIE	1740
3887	QY	-----	3886
1741	Db	DVKNOEAREGATAVLQCELSKAAPVEMWKGSETLRGGDRYSLRDQGTCELOIHGLSVAD	1800
3887	QY	-----	3886
1801	Db	TGEYSCVCGQERTSATLTVRALPARFTQDLTKYKASEGATATLQCELSKVAPEVWKGPE	1860
3887	QY	-----	3886
1861	Db	TLRDGGYSLRKQDGTCELOIHDLISVADAGEYSCMCGQERTSATLTVRDCHTLHWPHPY	1920
3887	QY	-----	3886
1921	Db	FQLPGLLKEPETLVIQIPSPVILFTEGLRNEEAMEGATATLQCELSKAAPVEMWKGLE	1980
3887	QY	-----	3886
1981	Db	ALRDGDYSLRDQGVACELQIHGLAMADNGYSSLPARFIEDMRNOKATEGATVTLQCKL	2040
3887	QY	-----	3886

Db	2041	RKAAPVWRKGPNTLKGDRYSLSKDGTSCEQLQIRGLVIADAGEYSCICEQERTSATLTV	2100
Qy	3887	-----	3886
Db	2101	RALPARPIEDVRNHEATEGATAVLQCELSKAAPVWRKGSSETLRDGRVLSLRDGTGRCEL	2160
Qy	3887	-----	3886
Db	2161	QIRGLAVEDTGYLVCVCGQERTSATLTVRALPARFIDNMTNOEAREGATATLHCELSKVA	2220
Qy	3887	-----	3886
Db	2221	PVWRKGPETLRDGRHSLSLRDGTGRCELQIRGLSVADAGEYSCVCGQERTSATLTVRALP	2280
Qy	3887	-----	3886
Db	2281	AKFTKGLRNEEATEGATAMLQCELSKVAPVWRKGPETLRDGRVNLRLDGTGRCELQIHG	2340
Qy	3887	-----	3886
Db	2341	LSVADTGEYSCVCGQERTSATLTVKAPQPVFRPELQSLQAEBSSTATLOCELSEPTATVV	2400
Qy	3922	WSKGGIQLQANGRRERPLQCTAELVLQDLQREDTGEYTCCTCGSQATSATLTVTAAPVRF	3981
Db	2401	WSKGGIQLQANGRRERPLQCTAELVLQDLQREDTGEYTCCTCGSQATSATLTVTAAPVRF	2460
Qy	3982	LRELQHQEVDDEGTAHLCCELSPAGASVWRKGSLOLFFCAKYQMVQDGAEEALLVRGVE	4041
Db	2461	LRELQHQEVDDEGTAHLCCELSPAGASVWRKGSLOLFFCAKYQMVQDGAEEALLVRGVE	2520
Qy	4042	QEDAGDYTCDTGHTQSMASLSVRVPRPKETRLQSLQEQETDIARLCCOLSDAESGAVVQ	4101
Db	2521	QEDAGDYTCDTGHTQSMASLSVRGGR-----	2549
Qy	4102	WLKEGVELHAGPKYENRSGOGATRELLIHOLEAKDTGEYACVTGGQKTAASLRVTPEVTI	4161
Db	2550	-----CGPOVRDAAOGATRELLIHOLEAKDTGEYACVTGGQKTAASLRVTPEVTI	2600
Qy	4162	VRGLVDAEVTADDEVEFSCEVSRAGATGVQWCLQGLPLQSNVEVAVRDGRIHTRLKLG	4221
Db	2601	VRGLVDAEVTADDEVEFSCEVSRAGATGVQWCLQGLPLQSNVEVAVRDGRIHTRLKLG	2660
Qy	4222	VTPEADAGTVSFHLGNHASSAQLTVRAPEVTIILEPLQDVOLSEQGDASFQCLRSASGOBA	4281
Db	2661	VTPEADAGTVSFHLGNHASSAQLTVRAPEVTIILEPLQDVQLR-----	2701
Qy	4282	RWALGGVPLOANEMNDITVBOGTLHLTLHKVLTLEDAGTVSFHVGTCSSEALQKVT---	4337
Db	2702	-----GVPIQANEMNDITVBOGTLHLTLHKVLTLEDAGTVSFHVGTCSSEALQKVTAVP	2756
Qy	4338	-----	4337
Db	2757	CLVRGGLQNVDFAGEVATFSCEDGQPSAIVARDGIFHSLMLSLGLGVADSGTVIFRAGPLV	2816
Qy	4338	-----	4337
Db	2817	STAKLLIKOPVVEVVSAMQDLAVEBGSSELLCOYSRPVQATWKMDEREVHTDGHVITIE	2876
Qy	4338	-----AKNTVVVRGLNVEALEGGEALPE	4360
Db	2877	QDMNVARLTFRPAIPCDSGIYSCEAAGTRVVALLVQVQAKNTVVVRGLNVEALEGGEALPE	2936
Qy	4361	COLSQSPEVAHNTWLLDDEPVRTSENAEVVFFENGRLHLLLLKNLRPQDSCRVTFIAGDMV	4420
Db	2937	COLSQSPEVAHNTWLLDDEPVRTSENAEVVFFENGRLHLLLLKNLRPQDSCRVTFIAGDMV	2996
Qy	4421	TSAFLTVR-----GWLEIILEPLKNAAVRAGAQAFTCTLSBAVPVGEASWINGAA	4472
Db	2997	TSAFLTVRGDCAVLVQGWLEIILEPLKNAAVRAGAQAFTCTLSBAVPVGEASWINGAA	3056
Qy	4473	VQPDSDSWTVTADGSHQALLLSAQPHHAGEVTFACRDVASARLTVLGLPDPDEAVVV	4532



Db 3057 VQDDSDWTVTADGSHHALLRSQAQPHHAGEVTFACRDVAVASARLTVLGLPDPEDAENV 3116  
Qy 4533 AHSHTVTLSWAAPMSGGGLCYRVEVEKAGATQWRLCHELVPGPECVWDGLAGETY 4592  
Db 3117 ARSHTVTLSWAAPMSGGGLCYRVEVEKAGATQWRLCHELVPGPECVWDGLAGETY 3176  
Qy 4593 RFRVAAGVPGAGBPVHLPTQVRLAEPPKVPPOPSAPESQVAAAGDVLSLELVAAAG 4652  
Db 3177 RFRVAAGVPGAGBPVHLPTQVRLAEPPKVPPOPSAPESQVAAAGDVLSLELVAAAG 3235  
Qy 4653 EVIWHKGMERTOPGGRFEVVSQGRQMLVIKFTAEQDGEYHCGLAGS:CPAAATFOVA 4712  
Db 3236 EVIWHKGMERTOPGGRFEVVSQGRQMLVIKFTAEQDGEYHCGLAGS:CPAAATFOVA 3295  
Qy 4713 LSPASVDEAPQSLPPBAAGQDGLHLLWEALARKRMSRPTLDSISELPEEDGORSORLP 4772  
Db 3296 LSPASVDEAPQSLPPBAAGQDGLHLLWEALARKRMSRPTLDSISELPEEDGORSORLP 3355  
Qy 4773 QEAEEVAPDLSEGYSTADELARTGDADLSHTSSDDESRAGTSPSLVTLKXAGRPGTSPLA 4832  
Db 3356 QEAEEVAPDLSEGYSTADELARTGDADLSHTSSDDESRAGTSPSLVTLKXAGRPGTSPLA 3415  
Qy 4833 SKVGAP-----AAPSVKQQOQEPL 4852  
Db 3416 SKVSFPNLACKERTFTPRAGRSLLGFVGADPAFFGSEARSACRTRCAAPPPRESLKREP- 3474  
Qy 4853 AAVRPPLGLDSTKDLG-----DPSMD-----KAAVKIQ-----AAPK 4884  
Db 3475 ASCLP--GAMEAVELARKLOBEATCSICLDYFTDPVMTTCGHNPCFACIQLSWEKARGKK 3532  
Qy 4885 GYKVRK-----EMKQSGPMFSHTFGTEAQVGDALRECVKASKADVRARWLKOG 4935  
Db 3533 GRRKRGSGFPCREKMSQPNLLFN-----RLTKVAEMA----- 3568  
Qy 4936 VELTDGRHHHDQLDGDTCSLLIAGLDADAGCYTCOVSNK-----FGVTHSACVVVSGS 4991  
Db 3569 -----QOH-----FOLQKD-----LCQEHHEPLKLFCKQDSPICVVCRE 3604  
Qy 4992 ESEAESSGGELDDAFRAARLRHLRFTKSPAEVSDBELFLSADEG----- 5038  
Db 3605 SRE-----HRLHRVL-----PAEEAVQGYKLEEDMEYLRQITRTGN 3643  
Qy 5039 --PAPEPPADHQ-TYREDEHICIRREAL-----TEARQAVTRFOEMF 5079  
Db 3644 LOAREEOSLAEMQGVKERRERI VLEFEKMNLVLEEQRLLQALETEEBETASRLRESV 3703  
Qy 5080 ATL---GTGVEIKLVE-----QGPRR-----VEMCISK-- 5104  
Db 3704 ACLDRQGSLELLLLQLLEERTQGLQMLQDMKEPLSRAALLVLIHGMNLVEFPVVSLLP 3763  
Qy 5105 -----ETPAVVPPEPLPSLLTSDAAP-----VFLTE 5131  
Db 3764 SPLVLIATKAHTQLGPGTFTDPDCEPTPLRISPE-PRPS--TEDVWPDATSAYPYLLLYE 3820  
Qy 5132 LQNGEVQDGYP-----VSFDCVVTQPMPSVRWPKDGKLLBEDDHMYNEDQOG 5180  
Db 3821 SRQRYLSSSEGGSCFKDRFVAPCAV-QQTA-----FSSGR-----HYWEVGNIT 3868  
Qy 5181 GHQLIITAVPADMG-----VYRCIAENSMGVSTKAEALRVDLTSTDYDTAADATESSYF 5236  
Db 3869 GDALWALGVCRDNVSRKDRPKC-PENGFWV-----VQLSK-----GTYLSTF 3911  
Qy 5237 SAQGYLSRE---QEGTESTDEGLPOVVEELRDLQVAPGTRLAKQLKVKGYGAPRPLY 5293  
Db 3912 SALTVMLEPPSHNGIFLDFEAG-----EVSFYSVSDGSHLHTYSQAT--FPGLQOP 3962  
Qy 5294 WFKDQCPQTAGAHIMRTGKILHTLEIISVTRDSQVAAAYISNAMGAAYASARLLVRGP 5353  
Db 3963 PFCLGAP-----KSGQWISTVTMAGV-KDLATRTGAVVTPALGA----- 4001  
Qy 5354 DEPEEKPADVHQLVPPRMILERFTPKKVKGSSITFSKVEGPRVPTVHMLREAEGRV 5413  
Db 4002 -----YAPSATETQS-----PAP-----MSPPAPE--- 4021

Qy 5414 LNIQDPTGTYTVASSAQOQSHLVLLDVGROHQGYTTCIASNAAGQALCSASLHVSLPKVE 5473  
Db 4022 -----BEHPG--VPSLAPRSA-----RACAAAPGSPSPRAA 4051  
Qy 5474 EOEKYEKALISTFLOGTTQALISAQGLSETASFADLGGQRKEEPLAAAKEALGHLSLAEVOTE 5533  
Db 4052 EAARRPADSTAFPL-SVRMAA-----PDLSTNLQEEATCA----- 4087  
Qy 5534 EFLQKLTSTIEMVSAKITQAKLOVPGGSDSDSKTBSAS--PRHGRSRPSSSI----- 5585  
Db 4088 ICLDYFTDVTWDCGHNFCREICRCWQGPARTAPSAASCRRGTGPTARLWRPRW 4147  
Qy 5586 QESSSESEDGARGEIFDIYVVTADYLPGLAB-QDAITL-----REGQYVEVLDAAHPL 5638  
Db 4148 RGACTERRRRR-----VPAHREPLAFCGDELRLCAACERSGEH----- 4189  
Qy 5639 RWJ-----VRTPTKSSSRQG-----WSPAYLDRRLK-LSPE----- 5671  
Db 4190 -WAHRVGRCTRPTKTSRPLEAGTMAANETLLSGAKLEKSLHRLKQMQDALLFOAQDET 4248  
Qy 5672 ---WGAAEAPPEPGEAVSEDEYKARLSSVIOELLSSEQAFVEELOFLOSHHL---QHLE 5724  
Db 4249 CVLWQADEGGEQRQNVLEFE---RLRLLAEGGTAARAEAGEELKQSAHLAELE 4305  
Qy 5725 RCPHVPIAVAGOKA 5738  
Db 4306 R--PLPAACAGAAA 4317

## RESULT 7

US-10-093-463-74  
; Sequence 74, Application US/10093463  
; Publication No. US20030208039A1  
; GENERAL INFORMATION:  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Shenoy, Suresh  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Gusev, Vladimir  
; APPLICANT: Pochart, Pascal  
; APPLICANT: Zhong, Mei  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Mezes, Peter  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Casman, Stacie  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Li, Li  
; APPLICANT: Zerhusen, Bryan  
; APPLICANT: Tchernev, Vellizar  
; APPLICANT: Gangolli, Esha  
; APPLICANT: Vernet, Corine  
; APPLICANT: Pena, Carol  
; APPLICANT: Burgess, Catherine  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Gorman, Linda  
; APPLICANT: Spaderna, Steven  
; APPLICANT: Voss, Edward  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Anderson, David  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Miller, Charles  
; APPLICANT: Taupier, Raymond J. Jr.  
; TITLE OF INVENTION: No. US20030208039A1 Antibodies that Bind to Antigenic Polypept.  
; TITLE OF INVENTION: Encoding The Antigens, and Methods of Use.  
; FILE REFERENCES: 21402-290A (Cura 590AT)  
; CURRENT APPLICATION NUMBER: US/10/093,463  
; CURRENT FILING DATE: 2002-06-24  
; PRIOR APPLICATION NUMBER: 60/283,675  
; PRIOR FILING DATE: 2001-04-14  
; PRIOR APPLICATION NUMBER: 60/338,092

```
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 74
; LENGTH: 4675
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-093-463-74

Query Match      25.5%; Score 10506; DB 15; Length 4675;
Best Local Similarity 56.1%; Pred. No. 0;
Matches 2252; Conservative 45; Mismatches 129; Indels 1586; Gaps 35;

QY 2595 MPLYNDSPFHEISHKGRHTLVLSIQRADAGIVRASSLKYSTSARLEVRVKPVVFLKALD 2654
Db 1 MPLYNDSPFHEISHKGRHTLVLSIQRADAGIVRASSLKYSTSARLEVRVKPVVFLKALD 60

QY 2655 DLSAEEGRTLALQCEVSDPEAHVVRKDGVLQGPSDKYDFLHTAGTRGLVHVHDSPEADG 2714
Db 61 DLSAEEGRTLALQCEVSDPEAHVVRKDGVLQGPSDKYDFLHTAGTRGLVHVHDSPEADG 120

QY 2715 LYTCHVGSSETRARVRVHDLVHVGITKELKTMVELEGESCFECVLGSHESASDPAMTVGG 2774
Db 121 LYTCHVGSSETRARVRVHDLVHVGITKELKTMVELEGESCFECVLGSHESASDPAMTVGG 180

QY 2775 KTVGSSRFQATROGRKYILVREAPSDAGEVVFVSRGLTSKASLIVRERPAAI1KPLE 2834
Db 181 KTVGSSRFQATROGRKYILVREAPSDAGEVVFVSRGLTSKASLIVRERPAAI1KPLE 240

QY 2835 DOWVAPGDEVELCELGRAGTPVHVKDKRAIKRSOKYDVVCGTWMVLVIRGASLKDAG 2894
Db 241 DOWVAPGDEVELCELGRAGTPVHVKDKRAIKRSOKYDVVCGTWMVLVIRGASLKDAG 300

QY 2895 EYTCEVEASKSTASLHVEEKANCFTTELTNLQVEEKGTAVFTCKTEHPAATVTRKGLLE 2954
Db 301 EYTCEVEASKSTASLHVEEKANCFTTELTNLQVEEKGTAVFTCKTEHPAATVTRKGLLE 360

QY 2955 LRASGKHQPSOEGTLTLTTLTSALEKADSDTYTCDIGQAQSRALLVQGRRVHIIEDLV 3014
Db 361 LRASGKHQPSOEGTLTLTTLTSALEKADSDTYTCDIGQAQSRALLVQGRRVHIIEDLV 420
```

Db	1501	VEWRKGPNTLKDGRDRLSKDQGTSCELQIRGLVIADAGEYSCICEQERTSATLTVRALPA	1560
Qy	3887	-----	3886
Db	1561	RFIEDVRNHEATEGATAVLOCELSKAAPVWEVRKGSSETLRDGRDRLSKDQGTSCELQIRGL	1620
Qy	3887	-----	3886
Db	1621	AVEDTGEYLCVCGQERTSATLTVRALPARFIDNMNQAREGATATLHCELSKVPVWR	1680
Qy	3887	-----	3886
Db	1681	KGPETLRDGRHSRQDGRSCELOIRGLAVVDAGEYSCVCGQERTSATLTVRALPARFIE	1740
Qy	3887	-----	3886
Db	1741	DVKNQAREGATAVLOCELSKAAPVWEVRKGSSETLRDGRDRLSKDQGTSCELQIRGLSVAD	1800
Qy	3887	-----	3886
Db	1801	TGEYSCVCGQERTSATLTVRALPARFTQDKTYEASEGATATLOCELSKVPVWEVRKGP	1860
Qy	3887	-----	3886
Db	1861	TLRDGRYSLKQDGTSCELQIHDLVSADAGEYSCMCGQERTSATLTVRDCHTLVMPHY	1920
Qy	3887	-----	3886
Db	1921	FOLPGLKEPEETLIYIQIPSPVILFTBGLRNEAMEGATATLOCELSKAAPVWEVRKGL	1980
Qy	3887	-----	3886
Db	1981	ALRDGKYSLRQDGAVCLEQIHGLAMDNVYSSLPARFIEDMRNOKATEGATVTLQCKL	2040
Qy	3887	-----	3886
Db	2041	RKAAPVWEVRKGPNTLKDGRDRLSKDQGTSCELQIRGLVIADAGEYSCICEQERTSATLTV	2100
Qy	3887	-----	3886
Db	2101	RALPARFIEDVRNHEATEGATAVLOCELSKAAPVWEVRKGSSETLRDGRDRLSKDQGTSC	2160
Qy	3887	-----	3886
Db	2161	QIRGLAVEDTGEYLCVCGQERTSATLTVRALPARFIDNMNQAREGATATLHCELSKVA	2220
Qy	3887	-----	3886
Db	2221	PWEVRKGPETLRDGRHSRQDGTSCELQIRGLSVADAGEYSCVCGQERTSATLTVRALP	2280
Qy	3887	-----	3886
Db	2281	AKFTKGLRNEATEGATAMLOCELSKVAPEVRKGPETLRDGRDRLSKDQGTSCELQIRGH	2340
Qy	3887	-----APQVFRPEPLQSLQABEGSTATLOCELSSEPTATVV	3921
Db	2401	WSKGGLOQANGREPRLOQCTAELVLQDLQREDTGEYTCGSGATSATLTVTAAPVRF	2460
Qy	3982	LRELQHOEVEGGTAHLCCELSRAGASVEVRKGSLOLFFCAKQVMVQDGAALILVRGVE	4041
Db	2461	LRELQHOEVEGGTAHLCCELSRAGASVEVRKGSLOLFFCAKQVMVQDGAALILVRGVE	2520
Qy	4042	QEDAGDYTCDTGHTQSMASLVRPFRPKFTRLOSLEQETGJARLCCOLSDAESGAVVQ	4101
Db	2521	QEDAGDYTCDTGHTQSMASLVRGGR	2549
Qy	4102	WLKEGVELHAGPKVEMRSQATRELLIHOLEAKDGTGEYACVTGGOKTAASLRVTEPEVTI	4161

QY 4992 ESEAESSGGELDDAFRAARRLHRLFRKTPAEVSDBELFLSADG----- 5038  
 Db 3605 SRE-----HELHVL-----PAEAVQGYKLKEEDMEYLREQITGN 3643  
 QY 5039 --PAPPEPADWQ--TYRDEHIFICIRFEAL-----TEARQAVTRFOEMF 5079  
 Db 3644 LQAREEQSLAEWQGVKERRRIVLEFEKMNLYLVEEQRLLQALETEEETASRLRESV 3703  
 QY 5080 ATL---GIGVEIKLVE-----QGPRR-----VENCISK-- 5104  
 Db 3704 ACLDQSHSLLELLQLERSTQFPLQMDKKEPLSRAALLVLIHGMNLVEFPVWSLP 3763  
 QY 5105 -----ETPAPVVPPEPPLPSLLTSDAAP-----VFLTE 5131  
 Db 3764 SPLYLIAKHAHTQLQPGTFTFDEPCPTPLPSPP-PPPS--TEDWVPDATSAVPLYLLYE 3820  
 QY 5132 LQNEBVDGYP-----VSFDCVVTGQMPFSVRWFKDGKLLBEDDHY 5172  
 Db 3821 SRQRYLGSSPEGSGFCSDRFVAYPCAV-QGTA-----FSSGR-----HY 3860

## RESULT 8

US-10-476-397-4

; Sequence 4, Application US/10476397

; Publication No. US20040115687A1

; GENERAL INFORMATION:

; APPLICANT: YUE, Henry

; APPLICANT: LEE, Ernestine A.

; APPLICANT: DUGGAN, Brendan M.

; APPLICANT: THANGAVELU, Kavitha

; APPLICANT: HONCHELL, Cynthia D.

; APPLICANT: DING, Li

; APPLICANT: JACKSON, Jennifer L.

; APPLICANT: BAUGHN, Mariah R.

; APPLICANT: KALLICK, Deborah A.

; APPLICANT: LEE, Sally

; APPLICANT: WARREN, Bridget A.

; APPLICANT: XU, Yuming

; APPLICANT: TRAN, Uyen K.

; APPLICANT: LAL, Preeti G.

; APPLICANT: THORNTON, Michael B.

; APPLICANT: HAFALIA, April J.A.

; APPLICANT: YAO, Monique G.

; APPLICANT: NGUYEN, Dannel B.

; APPLICANT: GANDHI, Ameena R.

; APPLICANT: KHAN, Farrah A.

; APPLICANT: CHAWLA, Narinder K.

; APPLICANT: GRIFFIN, Jennifer A.

; APPLICANT: CHINN, Anna M.

; APPLICANT: ELIOTT, Vicki S.

; APPLICANT: RAMKUMAR, Javalaxmi

; APPLICANT: ARVIZU, Chandra S.

; APPLICANT: FORSYTHE, Ian J.

; TITLE OF INVENTION: CELL ADHESION AND EXTRACELLULAR MATRIX PROTEINS

; FILE REFERENCE: PR-0968 USN

; CURRENT APPLICATION NUMBER: US/10/476,397

; CURRENT FILING DATE: 2003-10-31

; PRIOR APPLICATION NUMBER: PCT/US02/13874

; PRIOR FILING DATE: 2002-05-01

; PRIOR APPLICATION NUMBER: US 60/298,290

; PRIOR FILING DATE: 2001-05-02

; PRIOR APPLICATION NUMBER: US 60/292,468

; PRIOR FILING DATE: 2001-05-21

; PRIOR APPLICATION NUMBER: US 60/298,616

; PRIOR FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 60/301,672

; PRIOR FILING DATE: 2001-06-28

; PRIOR APPLICATION NUMBER: US 60/345,008

; PRIOR FILING DATE: 2002-01-04

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PERL Program

; SEQ ID NO 4

; LENGTH: 2328  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; OTHER INFORMATION: Incyte ID No: 7326129CD1  
 US-10-476-397-4

Query Match 23.8%; Score 9834.5; DB 16; Length 2328;  
 Best Local Similarity 84.6%; Pred. No. 0;  
 Matches 1941; Conservative 76; Mismatches 213; Indels 65; Gaps 8;

QY	2096	MEVQLSHADVDSWTRDGLRFQCGPTCHLAVRGPMTHTLTSLGRPEDSGLMVFKAEGVHT	2155
Db	1	MEVQLSHADVDSWTRDGLRFQCGPTCHLAVRGPMTHTLTSLGRPEDSGLMVFKAEGVHT	60
QY	2156	SARLVVTELPVSFSPRLQDVVTEKEKVTLECELSRPNDVRWLKDGVELRAGKTMATAA	2215
Db	61	SARLVVTELPVSFSPRLQDVVTEKEKVTLECELSRPNDVRWLKDGVELRAGKTMATAA	120
QY	2216	QGACRSLTIRCFADQGVVCDADHAOSSASVKVQ-----	2251
Db	121	QGACRSLTIRCFADQGVVCDADHAOSSASVKVQ-----	180
QY	2252	-----GRTYTLIYRVLAEADAGSIQFVAENAESRA 2281	
Db	181	CEVSHDEVPGWFEGSKLRPTDNVIRQEGRTYTLIYRVLAEADAGSIQFVAENAESRA 240	
QY	2282	QLRVKELPVTLRPLRDKIAMKGRGVLECOVSASAOVRWFKGSOELQPGPKYELVSDG	2341
Db	241	QLRVKELPVTLRPLRDKIAMKGRGVLECOVSASAOVRWFKGSOELQPGPKYELVSDG	300
QY	2342	LYRKLIISDVHAEDEDTYTCADGVKTSAQFFVEEQSITIVRGLQDVTWMPAPAFEECE	2401
Db	301	LYRKLIISDVHAEDEDTYTCADGVKTSAQFFVEEQSITIVRGLQDVTWMPAPAFEECE	360
QY	2402	TSIPSPRPKMLLGTVLQAGNVLEBOEGVTHMLRRTCTMTGTPVHFTVGRSRSAR	2461
Db	361	TSIPSPRPKMLLGTVLQAGNVLEBOEGVTHMLRRTCTMTGTPVHFTVGRSRSAR	420
QY	2462	LWVSDIPVLTPLRPEKKTGRELSQSVLSCDPRPAPKAVQWKDDTPLSPSEKFKMSLEGQ	2521
Db	421	LWVSDIPVLTPLRPEKKTGRELSQSVLSCDPRPAPKAVQWKDDTPLSPSEKFKMSLEGQ	480
QY	2522	MAELRIILMPADAGVRCQAGSAHSSTEVTVAREVTVTGPLQDAATBEGWASFCECL	2581
Db	481	MAELRIILMPADAGVRCQAGSAHSSTEVTVAREVTVTGPLQDAATBEGWASFCECL	540
QY	2582	SHEDEEVEWSLNGMPLYNDSFHEISHKGRHTLVKSIQRADAGIVRASSLKVSTSAELE	2641
Db	541	SHEDEEVEWSLNGMPLYNDSFHEISHKGRHTLVKSIQRADAGIVRASSLKVSTSAELE	600
QY	2642	VRVPVFLKALDLSAEEROTLALQCEVSDPEAHVVRWKDGVOLGSPDKYDFLHTAGTR	2701
Db	601	VRVPVFLKALDLSAEEROTLALQCEVSDPEAHVVRWKDGVOLGSPDKYDFLHTAGTR	660
QY	2702	GLVVHDSVPEDAGLYTCHVGSEETRAVRVHDLHVGITTKLKTMEVLEGSCSPCECVLSH	2761
Db	661	GLVVHDSVPEDAGLYTCHVGSEETRAVRVHDLHVGITTKLKTMEVLEGSCSPCECVLSH	720
QY	2762	ESADPAMWTVGKTVGSSSRFQATROGRKYILVVRBAAPSDAGEVVFVSRGLTSKASLI	2821
Db	721	ESADPAMWTVGKTVGSSSRFQATROGRKYILVVRBAAPSDAGEVVFVSRGLTSKASLI	780
QY	2822	VRERPAALIKPLEDQWVAPGEDVELRCELBRAGTPVHMLKDKRAIKRSOKYDVVCEGTMA	2881
Db	781	VRERPAALIKPLEDQWVAPGEDVELRCELBRAGTPVHMLKDKRAIKRSOKYDVVCEGTMA	840
QY	2882	MLVIRGASLKDAGEYTCVEEASKSTASLHVEKANCFTTEELTNLQVEEKGTAVFTCKTEH	2941
Db	841	MLVIRGASLKDAGEYTCVEEASKSTASLHVEKANCFTTEELTNLQVEEKGTAVFTCKTEH	900
QY	2942	PAATVTRKGLLELRASGKHQPSQEGTLRLTISALEKADSDTYTCDIGQAQSRQALLVQ	3001

Db 901 PAATVTVWRKGLLELRASGKHPQSBGLTLLTLLTSALEKADSDTYTCDIGQAQRAQLLVQ 960  
Qy 3002 GRRVHIIEDELDVVDVQSSATFRCRISPANYEPVHWFLLDKTPLHANELNEIDAQPGGYH 3061  
Db 961 GRRVHIIEDELDVVDVQSSATFRCRISPANYEPVHWFLLDKTPLHANELNEIDAQPGGYH 1020  
Qy 3062 VLTLRQALXDSGTIYFAGDQASAAALRVTEKPSVPSRELTATITEGEDTLVLCETST 3121  
Db 1021 VLTLRQALXDSGTIYFAGDQASAAALRVTEKPSVPSRELTATITEGEDTLVLCETST 1080  
Qy 3122 CDIPMCWTGKGTLRGSRARCOLSHEGHRALLTGATLQDSGRYKCEAGACSSSVRVH 3181  
Db 1081 CDIPVCMWTGKGTLRGSRARCOLSHEGHRALLTGATLQDSGRYKCEAGACSSSVRVH 1140  
Qy 3182 ARPVRFOEALKDLEVLBEGGAATRCVLSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLEL 3241  
Db 1141 ARPVRFOEALKDLEVLBEGGAATRCVLSSVAAPVKWCYGNVLRPGDKYSLRQEGAMLEL 1200  
Qy 3242 VYRNLRPQDSGRYSCSGDQTTATLTVTALPAQFICKLRNKEATGATATLCELKSKTA 3301  
Db 1201 VYRNLRPQDSGRYSCSGDQTTATLTVTALPAQFICKLRNKEATGATATLCELKSKAA 1260  
Qy 3302 PVWRKGSSETLRDGRYCLRDGAMCELQIRGLAMVDAAYSCVCGEERTSASLTIRPMP 3361  
Db 1261 PVWRKGSSETLRDGRYCLRDGAMCELQIRGLAMVDAAYSCVCGEERTSASLTIRPMP 1320  
Qy 3362 AHTGRLRHQBSIEGATATLCELKSKAAPVWRKGRSLRDGDRHSRLRDGAVCELOICG 3421  
Db 1321 AHTGRLRHQBSIEGATATLCELKSKAAPVWRKGRSLRDGDRHSRLRDGAVCELOICG 1380  
Qy 3422 LAVADAGSYSCVCGEERTSATLTVKALPAKTEGRLNEEAVEGATAMLWCELSKAPVAV 3481  
Db 1381 LAVADAGSYSCVCGEERTSATLTVKALPAKTEGRLNEEAVEGATAMLWCELSKAPVAV 1440  
Qy 3482 RKGPENLRDGRYTLRQGTCEIQLCGLAMADAGEYLCVCGQERTSATLTIRALPARFI 3541  
Db 1441 RKGPENLRDGRYTLRQGTCEIQLCGLAMADAGEYLCVCGQERTSATLTIRALPARFI 1500  
Qy 3542 EDVKNOAREGATVLOCELNSAAPVWRKGSSETLRDGRYSLRDQGTKCELOIRGLAMA 3601  
Db 1501 EDVKNOAREGATVLOCELNSAAPVWRKGSSETLRDGRYSLRDQGTKCELOIRGLAMA 1560  
Qy 3602 DTGEYSVCVCGERTSAMLTVRALPIKFTGELRNEEBATEGATVLRCELSKAPVWRKKGH 3661  
Db 1561 DTGEYSVCVCGERTSAMLTVRALPIKFTGELRNEEBATEGATVLRCELSKAPVWRKKGH 1620  
Qy 3662 ETLRDGRHSLRDQGARCELQIRGLVAEDAGEYLCMCKERTSAMLTVRAMPSKFIELR 3721  
Db 1621 ETLRDGRHSLRDQGARCELQIRGLVAEDAGEYLCMCKERTSAMLTVRAMPSKFIELR 1680  
Qy 3722 NEEATEGDTATLWCELSKAPVWRKKGHETLRDGRHSLRDQGRCELOIRGLAVVDAGE 3781  
Db 1681 NEEATEGDTATLWCELSKAPVWRKKGHETLRDGRHSLRDQGRCELOIRGLAVVDAGE 1740  
Qy 3782 YSCVCGQERTSATLTVRALPARFIEDVKNOAREGATVLOCELKSKAAPVWRKGSSETLR 3841  
Db 1741 YSCVCGQERTSATLTVRALPARFIEDVKNOAREGATVLOCELKSKAAPVWRKGSSETLR 1800  
Qy 3842 GGRYSLRDQGTCELOIHLGSLVADTGEYSVCVCGERTSATLTVRAPQVPRREPLOSQA 3901  
Db 1801 GGRYSLRDQGTCELOIHLGSLVADTGEYSVCVCGERTSATLTVRALPARFQDLKTKEA 1860  
Qy 3902 EBGSTATLOCELSSEPTATVNSKGLQLOANGREPRPLOGCTPAELVLQDQREDTGEYTC 3961  
Db 1861 SEGATATLOCELSK-VAPVWRKGPETLRDGRYSLKQDGTCELOIHLDSVADAGEYSC 1919  
Qy 3962 TCGSQATSATLTVTAAVPRFELQHOEVDGEGTAHLCCELSRAGASVWRKGSJQLPFC 4021  
Db 1920 MCCQERTSAMLTVRALPARFTEGLRNEEAMEGATATLOCELSKA-APVWRKGLBALRDG 1978  
Qy 4022 AKYQVQDGAABELLVRGVEQEDAGDYTCDTGHTOSMASLSVRPRPKFKTLQSLQEQET 4081

Db 1979 DKYSLRQDGAVCELOIHLGLAMADNGVYSCVCGQERTSATLTVRALPARFIEDMRNQKATE 2038  
Qy 4082 GDIAELCOLSDAESGAVQWMLKEGVELHAGPKYEMRSGGATRELLHLHLEAKDGTGEYAC 4141  
Db 2039 GATVTLQCKLRKA---APVWRKGNENTLKGDRYSLKQDGTSCELQIRGLVIADAGEYSC 2095  
Qy 4142 VTGGOKTAASLRVTEPEVTIIVRGLVDAEVTADEDEVEFSCEVSRAGATGVQMCLOGLPLOS 4201  
Db 2096 ICEQERTSATLTVRALPARFIEDVENHEATEGATVLOCELKSKAAP---VWRKGSSETLRD 2153  
Qy 4202 NEVTEVAVRDGRHITRLKGVTPPEDAGTVSFHLGNHASSAQLTVRAPEVTILEPQDVLQ 4261  
Db 2154 GDRYSILR-QDGTFRCELQIRGLAVEDTGYLCVCGQERTSATLTVRALPARFIDNNMQEA 2212  
Qy 4262 SSGQASFOCLSRASGOEARWALGCVPLQANEMNDITVEOQTLHLLTLHKVTLLEDAGTV 4321  
Db 2213 REGATATLCELKSKAPVE---WRKGRSLRDGDRHSLR-QDGAVCELOICGLAVADAGEY 2269  
Qy 4322 SFHVGTCSSEAQLKV 4336  
Db 2270 SCVCGEERTSATLTV 2284

RESULT 9  
US-09-858-664A-2  
; Sequence 2..Application US/09858664A  
; Patent No. US20020072491A1  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui, et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL000927-CIP  
; CURRENT APPLICATION NUMBER: US/09/858,664A  
; CURRENT FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 09/711,134  
; PRIOR FILING DATE: 2000-11-11  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 1665  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-858-664A-2

Query Match 20.4%; Score 8423; DB 9; Length 1665;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1612; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6356 PPSMQVTIEDVQAGTGTAGTQFEALIEGDPQPSVTWYKDSVQLVDSTRLSQOQEGTTSYLV 6415  
Db 53 PPSMQVTIEDVQAGTGTAGTQFEALIEGDPQPSVTWYKDSVQLVDSTRLSQOQEGTTSYLV 112  
Qy 6416 LRHVASKDAGVYTCQAQNTGGQVLCCKAELLVLDGDNPESEKSHRRKLSHSFYEVKEEIG 6475  
Db 113 LRHVASKDAGVYTCQAQNTGGQVLCCKAELLVLDGDNPESEKSHRRKLSHSFYEVKEEIG 172  
Qy 6476 RGVPFGFKRVQHKGNKILCAAKFIPLRSRTAQAYBERDILAALSHPLVTGLLDQPETRK 6535  
Db 173 RGVPFGFKRVQHKGNKILCAAKFIPLRSRTAQAYBERDILAALSHPLVTGLLDQPETRK 232  
Qy 6536 TLILILELCSSEELLDRLYRKGVVTEAEVRYVYIQQVLEGLHYLHSHGVHLHDIKPSNIML 6595  
Db 233 TLILILELCSSEELLDRLYRKGVVTEAEVRYVYIQQVLEGLHYLHSHGVHLHDIKPSNIML 292  
Qy 6596 VHPAREDIKCDGFAQNTIPASLOFSQYSGSPFVSPETIQQNPVSEASDIWANGVTSYL 6655  
Db 293 VHPAREDIKCDGFAQNTIPASLOFSQYSGSPFVSPETIQQNPVSEASDIWANGVTSYL 352  
Qy 6656 SLTCCSPFAGESDRATLLNVLEGRVSWSPMAHLSEDAKDFIKATIQRAFPARPSAAQC 6715  
Db 353 SLTCCSPFAGESDRATLLNVLEGRVSWSPMAHLSEDAKDFIKATIQRAFPARPSAAQC 412

QY 6716 LSHPWFLKSMPEAEAHFINTKQLKELLARSWORSIMSYKSTLYMRSIPPELLRPPDPS 6775  
Db 413 LSHPWFLKSMPEAEAHFINTKQLKELLARSWORSIMSYKSTLYMRSIPPELLRPPDPS 472  
QY 6776 LGVAREHLCRDTCGSS 6835  
Db 473 LGVAREHLCRDTCGSS 532  
QY 6836 ASASRSRTEAPAPASPGAGPAPAGGCPVPRHSVIRSLFYHQAGESPEHGALAGSRHP 6895  
Db 533 ASASRSRTEAPAPASPGAGPAPAGGCPVPRHSVIRSLFYHQAGESPEHGALAGSRHP 592  
QY 6896 ARRRHLLKGGYIAGALPGLRPLMEHVRLEBEAAREEQATLLAKAPSFETALRUPASGTH 6955  
Db 593 ARRRHLLKGGYIAGALPGLRPLMEHVRLEBEAAREEQATLLAKAPSFETALRUPASGTH 652  
QY 6956 LAPGSHSLEHDSPTPRPSSEACGEAQRPLPSAGGAPTRDMGHGPGSKOLPSTGGHPG 7015  
Db 653 LAPGSHSLEHDSPTPRPSSEACGEAQRPLPSAGGAPTRDMGHGPGSKOLPSTGGHPG 712  
QY 7016 TAQPERPSPDWGOPAPFCHPKQGSAPQEGCSPHPAVAPCPGSPFPGSCKEAPLPSS 7075  
Db 713 TAQPERPSPDWGOPAPFCHPKQGSAPQEGCSPHPAVAPCPGSPFPGSCKEAPLPSS 772  
QY 7076 PFLGQOPAPPAPAKASPLDSKMGPGDISLGRPKPGPCSPGASQASSQVSSLRVGS 7135  
Db 773 PFLGQOPAPPAPAKASPLDSKMGPGDISLGRPKPGPCSPGASQASSQVSSLRVGS 832  
QY 7136 SQVTEPGPSLDAECWTQEAEDLSDSTPTLQRPQEQATMRKFSLGGGGYAGVAGYTF 7195  
Db 833 SQVTEPGPSLDAECWTQEAEDLSDSTPTLQRPQEQATMRKFSLGGGGYAGVAGYTF 892  
QY 7196 FGGDAGMLGGPMMARIAMAVSSEEEQEEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 7255  
Db 893 FGGDAGMLGGPMMARIAMAVSSEEEQEEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 952  
QY 7256 EVGRAPTRSSPEPTPMDIGOVSLVOIRDLSGDAEAADTTSLDISEVDPAVNLSDLYDI 7315  
Db 953 EVGRAPTRSSPEPTPMDIGOVSLVOIRDLSGDAEAADTTSLDISEVDPAVNLSDLYDI 1012  
QY 7316 KYLPFFERMIKPKVKSQAQPPSPMAEEELAEPEPTPMPWPGELPHAGLEITEESD 7375  
Db 1013 KYLPFFERMIKPKVKSQAQPPSPMAEEELAEPEPTPMPWPGELPHAGLEITEESD 1072  
QY 7376 ALLAEAAVGRKRKWSPSRSLFHPGRLPLDPEAELGLBERKVASVEHISRLKGRPEG 7435  
Db 1073 ALLAEAAVGRKRKWSPSRSLFHPGRLPLDPEAELGLBERKVASVEHISRLKGRPEG 1132  
QY 7436 LEKEGPPKXPGLASPRLSGLKSWDRAPTFLRELSDTVLGSVTLACQVSAQPAQAAT 7495  
Db 1133 LEKEGPPKXPGLASPRLSGLKSWDRAPTFLRELSDTVLGSVTLACQVSAQPAQAAT 1192  
QY 7496 WSKDGAPLESRRVLIISATLKNFOLLTILVVAEDLGVYTCVSNALGTVTTTGVLRKAE 7555  
Db 1193 WSKDGAPLESRRVLIISATLKNFOLLTILVVAEDLGVYTCVSNALGTVTTTGVLRKAE 1252  
QY 7556 RPSSSPCPDIGEVYADGVLLVWKPVSIGPVTVIVOCLEGGSWTTLASDIFDCCVLTSK 7615  
Db 1253 RPSSSPCPDIGEVYADGVLLVWKPVSIGPVTVIVOCLEGGSWTTLASDIFDCCVLTSK 1312  
QY 7616 LSRGTTTFTTACVSKAGMGYSPSPQVLLGGPSHLASEEESQGRSAQFLPSTKTFAFQ 7675  
Db 1313 LSRGTTTFTTACVSKAGMGYSPSPQVLLGGPSHLASEEESQGRSAQFLPSTKTFAFQ 1372  
QY 7676 TQQRGRFSVVOCKEASGRALAAKIIIPHPKDKTAVLREVEALGLRPHLAQHAAY 7735  
Db 1373 TQQRGRFSVVOCKEASGRALAAKIIIPHPKDKTAVLREVEALGLRPHLAQHAAY 1432  
QY 7736 LSPRHVLVILELCSGPELLPCLAEASYSSEVKYLMQMLSATQYLNHQTLLHLDLRSE 7795  
Db 1433 LSPRHVLVILELCSGPELLPCLAEASYSSEVKYLMQMLSATQYLNHQTLLHLDLRSE 1492  
QY 7796 NMIIITEYNLLKVVDLGNASLSQEKVLPDSKFKDYLETWAPELLEGGQAVPQTDIWAIGV 7855

Db 1493 NMIIITEYNLLKVVDLGNASLSQEKVLPDSKFKDYLETWAPELLEGGQAVPQTDIWAIGV 1552  
QY 7856 TAFIMLSAEPYVSSEGARDLQRLKGLVRLSRCVAGLSGGAVAFURSTLCAQWGRPCA 7915  
Db 1553 TAFIMLSAEPYVSSEGARDLQRLKGLVRLSRCVAGLSGGAVAFURSTLCAQWGRPCA 1612  
QY 7916 SSCLOCPWLTGEGPACSRPAPVTFPTARLURVVRNREKRALLYKRHNLAQVR 7968  
Db 1613 SSCLOCPWLTGEGPACSRPAPVTFPTARLURVVRNREKRALLYKRHNLAQVR 1665

## RESULT 10

US-10-697-263-2  
; Sequence 2, Application US/10697263  
; Publication No. US20040063142A1  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui, et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEROPF  
; FILE REFERENCE: CLO00927-CIP-DIV2  
; CURRENT APPLICATION NUMBER: US/10/697,263  
; PRIOR FILING DATE: 2003-10-31  
; PRIOR APPLICATION NUMBER: 10/274,978  
; PRIOR FILING DATE: 2002-10-22  
; PRIOR APPLICATION NUMBER: 09/858,664  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 09/711,134  
; PRIOR FILING DATE: 2000-11-14  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 1665  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-697-263-2

Query Match 20.4%; Score 8423; DB 12; Length 1665;  
Best Local Similarity 99.9%; Pred. No. 0;

Matches 1612; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 6356 PSMQVTTIEDVQAQTGGTAQFEAIEGDPQPSVTWYKDSVOLVSTRLSQOQEGTYSLV 6415  
Db 53 PSMQVTTIEDVQAQTGGTAQFEAIEGDPQPSVTWYKDSVOLVSTRLSQOQEGTYSLV 112  
QY 6416 LRHVASKDAGVYTCQAQNTGGVLCVCKAEFLVLGGDNEPDSEKQSHRRKLHSFYEVKEIG 6475  
Db 113 LRHVASKDAGVYTCQAQNTGGVLCVCKAEFLVLGGDNEPDSEKQSHRRKLHSFYEVKEIG 172  
QY 6476 RGVFGVKKRVQHKNGKILCAAKFPLRSRTRAQYRERDILAAALSHPLVTGLDQFETR 6535  
Db 173 RGVFGVKKRVQHKNGKILCAAKFPLRSRTRAQYRERDILAAALSHPLVTGLDQFETR 232  
QY 6536 TLIIILELCSSEELLRLYKGVVTEAEVKYIQOOLVEGLYHSHGVHLHDIKPSNLM 6595  
Db 233 TLIIILELCSSEELLRLYKGVVTEAEVKYIQOOLVEGLYHSHGVHLHDIKPSNLM 292  
QY 6596 VHPAREDIKICDFGFAQNTITPAELQFSQYGSPEFVSPEIIQQNPVSEASDIWANGVISYL 6655  
Db 293 VHPAREDIKICDFGFAQNTITPAELQFSQYGSPEFVSPEIIQQNPVSEASDIWANGVISYL 352  
QY 6656 SLTCCSPFAGESDRATLNVLEGRVSWSPMAAHLSEDAKDFIKATLQAPAPPSAAQC 6715  
Db 353 SLTCCSPFAGESDRATLNVLEGRVSWSPMAAHLSEDAKDFIKATLQAPAPPSAAQC 412  
QY 6716 LSHPWFLKSMPEAEAHFINTKQLKELLARSWORSIMSYKSTLYMRSIPPELLRPPDPS 6775  
Db 413 LSHPWFLKSMPEAEAHFINTKQLKELLARSWORSIMSYKSTLYMRSIPPELLRPPDPS 472  
QY 6776 LGVAREHLCRDTCGSS 6835  
Db 473 LGVAREHLCRDTCGSS 532

QY	6836	AEASRSTEARPAPASPEGAGPAAQOCVPRHSVIRSLPYHOAGSPBHGALAPGSRHP	6895
DB	533	AEASRSTEARPAPASPEGAGPAAQOCVPRHSVIRSLPYHOAGSPBHGALAPGSRHP	592
QY	6896	ARRRHLKGGYIAGALPGLREPLMEHRVLEEEAAREEQATLLAKAPSPETALRULPASGTH	6955
DB	593	ARRRHLKGGYIAGALPGLREPLMEHRVLEEEAAREEQATLLAKAPSPETALRULPASGTH	652
QY	6956	LAPGHSHLEHDSPTPRPSSEACGGAORLPASPGGAPIRDMGHPOGSKOLPSTGGHPG	7015
DB	653	LAPGHSHLEHDSPTPRPSSEACGGAORLPASPGGAPIRDMGHPOGSKOLPSTGGHPG	712
QY	7016	TAQPERPSPDSFWQOAPFCHPKQSGAPOEGCSPHPAVAPCPPGSPFGSCKEAPLPVSS	7075
DB	713	TAQPERPSPDSFWQOAPFCHPKQSGAPOEGCSPHPAVAPCPPGSPFGSCKEAPLPVSS	772
QY	7076	PFLGQOPAPPAPAKASPLOSCKWPGDILSLGRPKPGCSPGSAQASQSSQVSSSLRVGS	7135
DB	773	PFLGQOPAPPAPAKASPLOSCKWPGDILSLGRPKPGCSPGSAQASQSSQVSSSLRVGS	8322
QY	7136	SOVGTEPSPLDASGWTQEAEDLS DSTPTLQRPQOATMRKPSLGGRGYAGVAGYTGA	7195
DB	833	SOVGTEPSPLDASGWTQEAEDLS DSTPTLQRPQOATMRKPSLGGRGYAGVAGYTGA	892
QY	7196	FGDAGGMLGQGPWWAIIANAVSQSBEEOQEARAESQSEEOQEARAESPLPQVSARPPV	7255
DB	893	FGDAGGMLGQGPWWAIIANAVSQSBEEOQEARAESQSEEOQEARAESPLPQVSARPPV	952
QY	7256	EYVRAPTRSSPEPTFWEDIQVSLVQIRDLSGDAEAA DTISLDISEVDPAYLNLSDIYDI	7315
DB	953	EYVRAPTRSSPEPTFWEDIQVSLVQIRDLSGDAEAA DTISLDISEVDPAYLNLSDIYDI	1012
QY	7316	KYLPPEFMI FRKVPKSAQPPPPSPMAEEELAEFPETWMPGELGPHAGLITESESDVD	7375
DB	1013	KYLPPEFMI FRKVPKSAQPPPPSPMAEEELAEFPETWMPGELGPHAGLITESESDVD	1072
QY	7376	ALLAEAAVGRKXKSSPSRSLFHPGGRHLPDSEPAELGIRREKVKASVEHISRLKGRPEG	7435
DB	1073	ALLAEAAVGRKXKSSPSRSLFHPGGRHLPDSEPAELGIRREKVKASVEHISRLKGRPEG	1132
QY	7436	LEKEGPPRKKPGLASFRLSLGKWDRAPTFLRELSDETIVLGQSVTLACQVSAQFAAQAT	7495
DB	1133	LEKEGPPRKKPGLASFRLSLGKWDRAPTFLRELSDETIVLGQSVTLACQVSAQFAAQAT	1192
QY	7496	NSKDGAPLESSRVLISATLKNFOLLTILVVAEDLGVTVCVSNALGTVTTTGVLRKAE	7555
DB	1193	NSKDGAPLESSRVLISATLKNFOLLTILVVAEDLGVTVCVSNALGTVTTTGVLRKAE	1252
QY	7556	RPSSPCPDIDIEVYADGVLLVWKVPBSYGPVTIYVQCSLEGGSWTTLASDIFDCYLTSK	7615
DB	1253	RPSSPCPDIDIEVYADGVLLVWKVPBSYGPVTIYVQCSLEGGSWTTLASDIFDCYLTSK	1312
QY	7616	LSRGCTYITFRACYSKAGMPYSSPQBQVLLGGPSHLAGEEBSQGRSAQPLPSTKTFAQ	7675
DB	1313	LSRGCTYITFRACYSKAGMPYSSPQBQVLLGGPSHLAGEEBSQGRSAQPLPSTKTFAQ	1372
QY	7676	TOIQRGRESVVRQCEWAKSGBALAAKII PPHKDKTAVLREYEALKGRHHPHLAQHAA	7735
DB	1373	TOIQRGRESVVRQCEWAKSGBALAAKII PPHKDKTAVLREYEALKGRHHPHLAQHAA	1432
QY	7736	LSPRHLVLI BELCGPELLPCLAERASYSESEYKDYLMQMLSATQYLNHQHILDLRSE	7795
DB	1433	LSPRHLVLI BELCGPELLPCLAERASYSESEYKDYLMQMLSATQYLNHQHILDLRSE	1492
QY	7796	NMIITTYNLLKVVDLGHQAQSLQSKVLPSDKFYDLETWAPELLBGQAGVQOTDIAIGV	7855
DB	1493	NMIITTYNLLKVVDLGHQAQSLQSKVLPSDKFYDLETWAPELLBGQAGVQOTDIAIGV	1552
QY	7856	TAFIMLSAEYPVSSSEGARDLQRLRGKLVRLSRCYAGLSGGAVAFRLSTLCAQPWGRPCA	7915
DB	1553	TAFIMLSAEYPVSSSEGARDLQRLRGKLVRLSRCYAGLSGGAVAFRLSTLCAQPWGRPCA	1612

Qy	7916	SSCLOCPWLTEGPACSRPAPVTFTPARLRVFNREKRALLYKRHNLAQVR	7968
D6	1613	SSCLOCPWLTEGPACSRPAPVTFTPARLRVFNREKRALLYKRHNLAQVR	1665

RESULT 11

```

US-10-415-011-22
; Sequence 22, Application US/10415011
; Publication No. US20040053394A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: XU, Yuming
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: YAO, Monique G.
; APPLICANT: RAMKOMAR, Jayalaxmi
; APPLICANT: DING, Li
; APPLICANT: TANG, Y. Tom
; APPLICANT: HAPALIA, April J.A.
; APPLICANT: NGUYEN, Dannel B.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: LU, Yan
; APPLICANT: YUE, Henry
; APPLICANT: BURFORD, Neil
; APPLICANT: BANDMAN, Olga
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: LAU, Preeti G.
; APPLICANT: RECIPON, Shirley A.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: BOROWSKI, Mark L.
; APPLICANT: THORNTON, Michael B.
; APPLICANT: SWARNAKER Anita
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: KHAN, Farrah A.
; APPLICANT: ISON, Craig H.
; TITLE OF INVENTION: HUMAN KINASES
; FILE REFERENCE: PI-0262 USN
; CURRENT APPLICATION NUMBER: US/10/415,011
; PRIOR FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: PCT/US01/47728
; PRIOR FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,410
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: US 60/244,068
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60/245,708
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/247,672
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/249,565
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/252,730
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: US 60/250,807
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040053394A1 7638121CD1
US-10-415-011-22

```

```
Query Match      20.4%; Score 8410; DB 12; Length 1665;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1610; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
; NAME/KEY: MISC_FEATURE
```



QY 6356 PPSMQVTIEDVQOTGGTAQFAALIEGDPQPSVWYKOSVOLVDSTRLSQQQEGTYSLV 6415  
Db 53 PPSMQVTIEDVQOTGGTAQFAALIEGDPQPSVWYKOSVOLVDSTRLSQQQEGTYSLV 112  
QY 6416 LRHVASKDAGVYTCIAQNTGGVLCVCAELLVLLGGNEPDSEKSHRKLHFSYFVEKEIG 6475  
Db 113 LRHVASKDAGVYTCIAQNTGGVLCVCAELLVLLGGNEPDSEKSHRKLHFSYFVEKEIG 172  
QY 6476 RGVFGRVQVQKGNKILCAAKFIPLRSTRQAAYRERDILAAALSHPLVTGLLOFETR 6535  
Db 173 RGVFGRVQVQKGNKILCAAKFIPLRSTRQAAYRERDILAAALSHPLVTGLLOFETR 232  
QY 6536 TLILILELCCSBEILDRLRYKQVTEAEVKYIQQLVEGLHSHGVHLHDIKPSNLM 6595  
Db 233 TLILILELCCSBEILDRLRYKQVTEAEVKYIQQLVEGLHSHGVHLHDIKPSNLM 292  
QY 6596 VHPAREDIKICDFGAQNTTAELOFSQYSGPEFVSPPIIQONPVSEASDIWAGVLSYL 6655  
Db 293 VHPAREDIKICDFGAQNTTAELOFSQYSGPEFVSPPIIQONPVSEASDIWAGVLSYL 352  
QY 6656 SLTCSPPAGSDRATLNLVLEGRVSWSPMAAHLSEDAKDFIKATLQAPQAPSAQC 6715  
Db 353 SLTCSPPAGSDRATLNLVLEGRVSWSPMAAHLSEDAKDFIKATLQAPQAPSAQC 412  
QY 6716 LSHPWFLKMAEAEAHFNTKQLFLLARSRWQSLMSYKSLVMSIPILLRPGFLRPSASLPEE 6835  
Db 413 LSHPWFLKMAEAEAHFNTKQLFLLARSRWQSLMSYKSLVMSIPILLRPGFLRPSASLPEE 532  
QY 6835 LGVAHLCRDTCGGSSSSSSSSSDNELAPPARAKSLPPSPVTHSPLLHPRGFLRPSASLPEE 6835  
Db 473 LGVAHLCRDTCGGSSSSSSSSSDNELAPPARAKSLPPSPVTHSPLLHPRGFLRPSASLPEE 532  
QY 6836 AEASERSTEAPASPAGPAPPAQCVPHSVIRSLFYHQAGESPEHGALAPGSRHP 6895  
Db 533 AEASERSTEAPASPAGPAPPAQCVPHSVIRSLFYHQAGESPEHGALAPGSRHP 592  
QY 6896 ARRHLLKGGYIAGALPGLREPLMEHRLVBEAEAREEQATLLAKAPSFETALRLPASGTH 6955  
Db 593 ARRHLLKGGYIAGALPGLREPLMEHRLVBEAEAREEQATLLAKAPSFETALRLPASGTH 652  
QY 6956 LAPGHSLSHDSSTPRPSEACEAORLPSPAGCAPTRDMGHGPGSKQLSTGCHPG 7015  
Db 653 LAPGHSLSHDSSTPRPSEACEAORLPSPAGCAPTRDMGHGPGSKQLSTGCHPG 712  
QY 7016 TAQPERFSPSPWQAPAFCHPKQGSAPQEGCSPPHAPVAPCPGSPPPGSCKEAPLVPS 7075  
Db 713 TAQPERFSPSPWQAPAFCHPKQGSAPQEGCSPPHAPVAPCPGSPPPGSCKEAPLVPS 772  
QY 7076 PFLGQOAPAPAKASPPLDSKMGPGDISLPGRKPGCSPGSSGASQAQSSQVSSLRVGS 7135  
Db 773 PFLGQOAPAPAKASPPLDSKMGPGDISLPGRKPGCSPGSSGASQAQSSQVSSLRVGS 832  
QY 7136 SQVTEPGSLDAEGWTOEADLSSTPTLQRPQEQATMEKSLGGRGVAGVAGTGA 7195  
Db 833 SQVTEPGSLDAEGWTOEADLSSTPTLQRPQEQATMEKSLGGRGVAGVAGTGA 892  
QY 7196 FGDAGMLCGGPMWARIAMVQSEEEEOEAEASQSEEQEAEASPLPQVSARVP 7255  
Db 893 FGDAGMLCGGPMWARIAMVQSEEEEOEAEASQSEEQEAEASPLPQVSARVP 952  
QY 7256 EVGRAPTRSPPTPVEDIQVSLVQIRDLSDGAEEADTISLDSIVDPAYLNLSDLYDI 7315  
Db 953 EVGRAPTRSPPTPVEDIQVSLVQIRDLSDGAEEADTISLDSIVDPAYLNLSDLYDI 1012  
QY 7316 KYLPFFFMIFRKYKSAQPPSPMAEEELAEFFPTWPGELGPHAGLEITESESDVD 7375  
Db 1013 KYLPFFFMIFRKYKSAQPPSPMAEEELAEFFPTWPGELGPHAGLEITESESDVD 1072  
QY 7376 ALLAEAAVGRKRWSSPSLSLHFCRHLPLDEPAELGLRERVKASVEHISRLKGRREG 7435  
Db 1073 ALLAEAAVGRKRWSSPSLSLHFCRHLPLDEPAELGLRERVKASVEHISRLKGRREG 1132  
QY 7436 LEKEGPPRKKPGLASPRLSGLKSWDRAPFTFLRELSDETVVLQSQVTLACQVSAQAQAT 7495

Db 1133 LEKEGPPRKKPGLASPRLSGLKSWDRAPFTFLRELSDETVVLQSQVTLACQVSAQAQAT 1192  
QY 7496 WSKDGAFLPSSSRVLISATLKNPQLTILVVAEDLVGVTCVSNALGTVTITGVLRLKAE 7555  
Db 1193 WSKDGAFLPSSSRVLISATLKNPQLTILVVAEDLVGVTCVSNALGTVTITGVLRLKAE 1252  
QY 7556 RPSSSPCPDIEGVYADGVLLVMKPVESYGPVTYIQQCSLEGSWTTLASDIFDCCVLSK 7615  
Db 1253 RPSSSPCPDIEGVYADGVLLVMKPVESYGPVTYIQQCSLEGSWTTLASDIFDCCVLSK 1312  
QY 7616 LSRGGTYTTRTACVSKAGMPYSSPSEQVLLGGPSHLASEEBSQGRSAQPLSTKTFAFO 7675  
Db 1313 LSRGGTYTTRTACVSKAGMPYSSPSEQVLLGGPSHLASEEBSQGRSAQPLSTKTFAFO 1372  
QY 7676 TQIORGRFVVRQCKEKASGRALAAKIIPYHPKDKTAVLREYEALKGLRHPHQAOLHAAY 7735  
Db 1373 TQIORGRFVVRQCKEKASGRALAAKIIPYHPKDKTAVLREYEALKGLRHPHQAOLHAAY 1432  
QY 7736 LSPRHLVLILELCSGPPELLPCLAEASYSSESEVKDYLMQMLSATQYLNQHLHLDLSE 7795  
Db 1433 LSPRHLVLILELCSGPPELLPCLAEASYSSESEVKDYLMQMLSATQYLNQHLHLDLSE 1492  
QY 7796 NMITEYNLLKVVDLGNQASLSQEKVLPSPDKFYLETMAPELLEGQGAVPQTDIWAIGV 7855  
Db 1493 NMITEYNLLKVVDLGNQASLSQEKVLPSPDKFYLETMAPELLEGQGAVPQTDIWAIGV 1552  
QY 7856 TAFIMLSAEYPVSSSGARDLQGLKGLVRLSRVAGLSGGAVFLRSTLCAQPMGRPCA 7915  
Db 1553 TAFIMLSAEYPVSSSGARDLQGLKGLVRLSRVAGLSGGAVFLRSTLCAQPMGRPCA 1612  
QY 7916 SSCLOCPMLTEBGPACSPAPVPTPTARLVRVVRNKRKALLYKRNHNAQVR 7968  
Db 1613 SSCLOCPMLTEBGPACSPAPVPTPTARLVRVVRNKRKALLYKRNHNAQVR 1665

## RESULT 12

US-10-182-243-46  
; Sequence 46, Application US/10182243  
; Publication No. US20040048310A1  
; GENERAL INFORMATION:  
; APPLICANT: FLOWMAN, GREGORY D.  
; APPLICANT: WHYTE, DAVID  
; APPLICANT: MANNING, GERARD  
; APPLICANT: SUDARSANAM, SUCHA  
; APPLICANT: MARTINEZ, RICARDO  
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES AND PROTEIN KINASE-LIKE  
; FILE REFERENCE: 038602/1366  
; CURRENT APPLICATION NUMBER: US/10/182,243  
; PRIOR FILING DATE: 2003-07-07  
; PRIOR APPLICATION NUMBER: PCT/US01/02337  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 46  
; LENGTH: 1618  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-182-243-46

Query Match 20.4%; Score 8407; DB 12; Length 1618;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1612; Conservative 0; Mismatches 0; Indels 6; Gaps 1;

QY 6357 PSMQVTIEDVQOTGGTAQFAALIEGDPQPSVWYKOSVOLVDSTRLSQQQEGTYSLV 6416  
Db 1 PSMQVTIEDVQOTGGTAQFAALIEGDPQPSVWYKOSVOLVDSTRLSQQQEGTYSLV 60  
QY 6417 RHVASKDAGVYTCIAQNTGGVLCVCAELLVLLGGNEPDSEKSHRKLHFSYFVE 6470  
Db 61 RHVASKDAGVYTCIAQNTGGVLCVCAELLVLLGGNEPDSEKSHRKLHFSYFVE 120





```

Db      241  AREDIKIDFGAQNITPAELQFSGYGFVSPFPIIQONFVSEASDIWAMGVSYLSLT 300
QY      6659  CSSPPAGESDRATLLNVLEGRVSMSPMAAHLSEDAKDFIKATLQRAPOARPSAAQCLSH 6718
Db      301  CSSPPAGESDRATLLNVLEGRVSMSPMAAHLSEDAKDFIKATLQRAPOARPSAAQCLSH 360
QY      6719  PWFILKSMFAEAEAHFINTKQLFLLARSRQWRLMSYKSIILVMSIPELLRPPSPSLGV 6778
Db      361  PWFILKSMFAEAEAHFINTKQLFLLARSRQWRLMSYKSIILVMSIPELLRPPSPSLGV 420
QY      6779  ARLHCRDTGGSSSSSSSDNELAPFARAKSLPPSPVTHSPILLHPRGFLRPSASIPERAEA 6838
Db      421  ARLHCRDTGGSSSSSSSDNELAPFARAKSLPPSPVTHSPILLHPRGFLRPSASIPERAEA 480
QY      6839  SERSTEAPAPASPEGAGPPAAQGCVPRHVSIRSLFYHQAGESPEHGALAPGSRHRPARR 6898
Db      481  SERSTEAPAPASPEGAGPPAAQGCVPRHVSIRSLFYHQAGESPEHGALAPGSRHRPARR 540
QY      6899  RHLKGGYIAGALPGLREPLMEHVRVLEBEAAREEQATLLAKAPSFETALPLPASGTHLAP 6958
Db      541  RHLKGGYIAGALPGLREPLMEHVRVLEBEAAREEQATLLAKAPSFETALPLPASGTHLAP 600
QY      6959  GHSHLEHDSPTTRPSSSEACGEAQLPSPAGGAPIRDMGHPOGSKQLPSTGHPGTAQ 7018
Db      601  GHSHLEHDSPTTRPSSSEACGEAQLPSPAGGAPIRDMGHPOGSKQLPSTGHPGTAQ 660
QY      7019  PERSPDPSWQGPAPFCHPKGPSARPOGCSBHPAVAPCPGSPPGCKEAPLVPSSPFL 7078
Db      661  PERSPDPSWQGPAPFCHPKGPSARPOGCSBHPAVAPCPGSPPGCKEAPLVPSSPFL 720
QY      7079  GQOAPAPAKASPLDLSKMGFGDISLPGRPKPCSPGASASQSSQVSSLRVSGSQV 7138
Db      721  GQOAPAPAKASPLDLSKMGFGDISLPGRPKPCSPGASASQSSQVSSLRVSGSQV 780
QY      7139  GTEPGSLDAGWTQEAEDLSDSPTTLQRPQEQATMRKFSLGGGGVAGVAGTFAFG 7198
Db      781  GTEPGSLDAGWTQEAEDLSDSPTTLQRPQEQATMRKFSLGGGGVAGVAGTFAFG 840
QY      7199  DAGMGLGCPMWARIANAVSSEEEQEARAESQSEEQEARAESPLPOVSARVPVEVG 7258
Db      841  DAGMGLGCPMWARIANAVSSEEEQEARAESQSEEQEARAESPLPOVSARVPVEVG 900
QY      7259  RAPTRSSPEPTPWEIDIGQVSLVQIRDLSDGAEAADTISLDISEVDPAYLNLSDLYIKYL 7318
Db      901  RAPTRSSPEPTPWEIDIGQVSLVQIRDLSDGAEAADTISLDISEVDPAYLNLSDLYIKYL 960
QY      7319  PFEFMIFRKVPKSAQPPSPMAEELAEPEPTMPWPGELGPHAGLEITEESEDVALL 7378
Db      961  PFEFMIFRKVPKSAQPPSPMAEELAEPEPTMPWPGELGPHAGLEITEESEDVALL 1020
QY      7379  AEAAGVGRKRWSSPSRSLFHPGSHLPLDEPAELGLRERVKASVEHISRLIKGRPEGLEK 7438
Db      1021  AEAAGVGRKRWSSPSRSLFHPGSHLPLDEPAELGLRERVKASVEHISRLIKGRPEGLEK 1080
QY      7439  EGPRKPKGLASPLSGIKSWDRAPTFLRELSDETIVLGOSVTLAQVSAQPAQAOTWTK 7498
Db      1081  EGPRKPKGLASPLSGIKSWDRAPTFLRELSDETIVLGOSVTLAQVSAQPAQAOTWTK 1140
QY      7499  DGAPLESSRSLVLSATLKNFOLLTILVVVAEDLGVTYCSVSNALGTVTITGLVKRKAERPS 7558
Db      1141  DGAPLESSRSLVLSATLKNFOLLTILVVVAEDLGVTYCSVSNALGTVTITGLVKRKAERPS 1200
QY      7559  SSPCPDICEVYADVOLLVWKPVEYSGVTIIVQCSLEGGSWTTLASDIFDCCYLTSLKR 7618
Db      1201  SSPCPDICEVYADVOLLVWKPVEYSGVTIIVQCSLEGGSWTTLASDIFDCCYLTSLKR 1260
QY      7619  GGTITFTACVSKAGMGPYSSQEOVLGGPSSHASEERSQGRSAOPLPSTKTFAPQTOI 7678
Db      1261  GGTITFTACVSKAGMGPYSSQEOVLGGPSSHASEERSQGRSAOPLPSTKTFAPQTOI 1320
QY      7679  QGRFVSVROCKEAKSRAALAAKIIPIYHPKDKTAVLREYALKGLRHPHQAOLHAAYLSP 7738

```

```

Db      1321  QGRFVSVROCKEAKSRAALAAKIIPIYHPKDKTAVLREYALKGLRHPHQAOLHAAYLSP 1380
QY      7739  RHLVLILELCSGPPELLPCLAEASYSSESVKDYLMQMLSATQYLNQHILHLDLSENNI 7798
Db      1381  RHLVLILELCSGPPELLPCLAEASYSSESVKDYLMQMLSATQYLNQHILHLDLSENNI 1440
QY      7799  ITEYNLLKVVVDLGNAGSLSQEKVLPSPDKPKDYLETWAPELLEGGQAVPQTDIWAIGVTAF 7858
Db      1441  ITEYNLLKVVVDLGNAGSLSQEKVLPSPDKPKDYLETWAPELLEGGQAVPQTDIWAIGVTAF 1500
QY      7859  IMLSABYPVSSSGARDLQRLKGLVRLSRVAGLSGGAVAFRLSTLCAQPKGRPCASSC 7918
Db      1501  IMLSABYPVSSSGARDLQRLKGLVRLSRVAGLSGGAVAFRLSTLCAQPKGRPCASSC 1560
QY      7919  LQCPWLTEEGPACSRPAPVTFTFARLVFVRNREXRALLYKRHNLAQVR 7968
Db      1561  LQCPWLTEEGPACSRPAPVTFTFARLVFVRNREXRALLYKRHNLAQVR 1610

```

## RESULT 14

```

US-10-307-019-1
; Sequence 1, Application US/10307019
; Publication No. US20030108533A1
; GENERAL INFORMATION:
; APPLICANT: Zeng Wenlin
; APPLICANT: Stanton, Lawrence
; APPLICANT: SciOS, INC.
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION
; FILE REFERENCE: SCIOS.021DV1
; CURRENT APPLICATION NUMBER: US/10/307,019
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/548,473
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/129,552
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1351
; TYPE: PAT
; ORGANISM: Homo sapiens
US-10-307-019-1

```

```

Query Match      17.1%; Score 7073; DB 14; Length 1351;
Best Local Similarity 99.9%; Pred. No. 3.6e-316;
Matches 1349; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      6619  LQFSQYGSPEFVSPHIIQONFVSEASDIWAMGVSYLSLTCSPPAGESDRATLLNVLEG 6678
Db      2    VQFSQYGSPEFVSPHIIQONFVSEASDIWAMGVSYLSLTCSPPAGESDRATLLNVLEG 61
QY      6679  RVSWSSPMAAHLSEDAKDFIKATLQRAPOARPSAAQCLSHPWFLKSPAEAEHFINTKQL 6738
Db      62    RVSWSSPMAAHLSEDAKDFIKATLQRAPOARPSAAQCLSHPWFLKSPAEAEHFINTKQL 121
QY      6739  KFLARSRQWRLMSYKSIILVMSIPELLRPPSPSLGVARHLCRDITGSSSSSSSDN 6798
Db      122  KFLARSRQWRLMSYKSIILVMSIPELLRPPSPSLGVARHLCRDITGSSSSSSSDN 181
QY      6799  ELAPPARAKSLPPSPVTHSPILLHPRGFLRPSASIPERAEASERSTEAPAPASPEGAGPP 6858
Db      182  ELAPPARAKSLPPSPVTHSPILLHPRGFLRPSASIPERAEASERSTEAPAPASPEGAGPP 241
QY      6859  AAQGCVPRHVSIRSLFYHQAGESPEHGALAPGSRHRPARRHLLKGGYIAGALPGLREPL 6918
Db      242  AAQGCVPRHVSIRSLFYHQAGESPEHGALAPGSRHRPARRHLLKGGYIAGALPGLREPL 301
QY      6919  MEHVRVLEBEAAREEQATLLAKAPSFETALPLPASGTHLAPCHSHLSEHDSPTTRPSSSEA 6978
Db      302  MEHVRVLEBEAAREEQATLLAKAPSFETALPLPASGTHLAPCHSHLSEHDSPTTRPSSSEA 361
QY      6979  CGEAQRLPSPASGGAPIRDMGHPOGSKQLPSTGHPGTAQPERPSPDPSWQGPAPFCHPK 7038

```

Db 362 CGEALRUPSPGCGAPIRDMGHGPGSKQLPSTGCHGPGTAQPERPSPDPSWGQPAFPCHPK 421  
Qy 7039 QGSAPOGSCSHPAVACPCSPGSCKEAPLVSSPFLGQCAPAPAKASPPPLDSKM 7098  
Db 422 QGSAPOGSCSHPAVACPCSPGSCKEAPLVSSPFLGQCAPAPAKASPPPLDSKM 481  
Qy 7099 GPGDISLPGREKPGPCSPGASQASSQVSSLRVSSQVGTGPGSLDAEGHTQEAEDL 7158  
Db 482 GPGDISLPGREKPGPCSPGASQASSQVSSLRVSSQVGTGPGSLDAEGHTQEAEDL 541  
Qy 7159 SDSTFTLQRPQOATMKRFSLGGGGGAGVAGYGTAFPGDAGMLGQGPWARIWAWS 7218  
Db 542 SDSTFTLQRPQOATMKRFSLGGGGGAGVAGYGTAFPGDAGMLGQGPWARIWAWS 601  
Qy 7219 QSEEEQEEAARAEQSEEQEAEARAEPLQVARSAPVEVGRAPTRSSPEPTPEDIQVQS 7278  
Db 602 QSEEEQEEAARAEQSEEQEAEARAEPLQVARSAPVEVGRAPTRSSPEPTPEDIQVQS 661  
Qy 7279 LVQIRDSGDABAAADTISLISEVDPAVNLSDLYDKYLPFFPMIRKVPKSAQPPPS 7338  
Db 662 LVQIRDSGDABAAADTISLISEVDPAVNLSDLYDKYLPFFPMIRKVPKSAQPPPS 721  
Qy 7339 PMAEEELAEFPPTWPGPELGPAGLEITESEDDVALLAAEAAVGRKRWSSPSRSLFH 7398  
Db 722 PMAEEELAEFPPTWPGPELGPAGLEITESEDDVALLAAEAAVGRKRWSSPSRSLFH 781  
Qy 7399 FPGRHLPLDEPAELGLRERVKASVEHISRIILKGRPEGLEKEGPPRKPGLASFLSLGKS 7458  
Db 782 FPGRHLPLDEPAELGLRERVKASVEHISRIILKGRPEGLEKEGPPRKPGLASFLSLGKS 841  
Qy 7459 WDRAPTFRLSDETVLQSVTLACQVSAQPAQAQATWSKDGAFLSSSRVLISATLKNF 7518  
Db 842 WDRAPTFRLSDETVLQSVTLACQVSAQPAQAQATWSKDGAFLSSSRVLISATLKNF 901  
Qy 7519 QLLTLVVAEDLGVYTCVSNALGTVTGTVLKAERPPSSPCPDIGEVYADGVLLVWK 7578  
Db 902 QLLTLVVAEDLGVYTCVSNALGTVTGTVLKAERPPSSPCPDIGEVYADGVLLVWK 961  
Qy 7579 PVESGPTVYIVQCSLEGGSWTLASDIFDCCYLTSLKSRGTYTFRACVSKAGMPYS 7638  
Db 962 PVESGPTVYIVQCSLEGGSWTLASDIFDCCYLTSLKSRGTYTFRACVSKAGMPYS 1021  
Qy 7639 SPSEQVLGGPSHLASEESQGRSAQPLPSTKTFATQIQGRFVSVRQWEXKASGRAL 7698  
Db 1022 SPSEQVLGGPSHLASEESQGRSAQPLPSTKTFATQIQGRFVSVRQWEXKASGRAL 1081  
Qy 7699 AAKIIPYHPKDTAVLREYALKGLRHPHLAQLHAAYLSRHLVLIILELCSGPPELLPCLA 7758  
Db 1082 AAKIIPYHPKDTAVLREYALKGLRHPHLAQLHAAYLSRHLVLIILELCSGPPELLPCLA 1141  
Qy 7759 ERASYSSEVKDYLWQMSATQYLNHNLHLDLRSENMIITEYNLLKVVLDLGNASLSQ 7818  
Db 1142 ERASYSSEVKDYLWQMSATQYLNHNLHLDLRSENMIITEYNLLKVVLDLGNASLSQ 1201  
Qy 7819 EXVLPDSKDFDYLETWAPELLEGGQAVPQTDIWAIGVTAFLMSAEYVPSSEGARDLQRG 7878  
Db 1202 EXVLPDSKDFDYLETWAPELLEGGQAVPQTDIWAIGVTAFLMSAEYVPSSEGARDLQRG 1261  
Qy 7879 LRKGLVRLSRCVAGLSGGVAFLRSTLCAQWGRPCASSCLOCPLWLTBEGPACSRPAPVT 7938  
Db 1262 LRKGLVRLSRCVAGLSGGVAFLRSTLCAQWGRPCASSCLOCPLWLTBEGPACSRPAPVT 1321  
Qy 7939 FPTARLVFVRNREKRALLYKRNHNAQVR 7968  
Db 1322 FPTARLVFVRNREKRALLYKRNHNAQVR 1351

RESULT 15  
US-10-408-765A-992  
; Sequence 992, Application US/10408765A  
; Publication No. US20040101874A1  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.

; APPLICANT: Fahy, Bojin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibsen, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Warnock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
; FILE REFERENCE: 660088.465  
; CURRENT APPLICATION NUMBER: US/10/408,765A  
; CURRENT FILING DATE: 2003-04-04  
; NUMBER OF SEQ ID NOS: 3077  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 992  
; LENGTH: 1596  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-408-765A-992

Query Match 16.0%; Score 6619.5; DB 16; Length 1596;  
Best Local Similarity 81.7%; Pred. No. 3,1e-295; Indels 21; Gaps 7;  
Matches 1309; Conservative 70; Mismatch 203;

Qy 2726 RARVRVHDLRVGTTKELKTMEVLEGECSFEVLSHESASDPAMTVGGKTVGSSSRFOA 2785  
Db 1 RARVRVHDLRVGTTKELKTMEVLEGECSFEVLSHESASDPAMTVGGKTVGSSSRFOA 60

Qy 2786 TROGRKIIVREAPSDAGEVVFVSRGLTSKASLIVRERPAALIKPLEDQWVAGEDVE 2845  
Db 61 TROGRKIIVREAPSDAGEVVFVSRGLTSKASLIVRERPAALIKPLEDQWVAGEDVE 120

Qy 2846 LRCELSRAGTPVHMLKDKAIRKSQYDVVCGTMAWLVRGASLKDAGEYTCVEBASKS 2905  
Db 121 LRCELSRAGTPVHMLKDKAIRKSQYDVVCGTMAWLVRGASLKDAGEYTCVEBASKS 180

Qy 2906 TASLHVEEKANCFTBELTNLQVEEKGTAVFTCKTEHPAATVTVRKGLLBELRASKGHPQSQ 2965  
Db 181 TASLHVEEKANCFTBELTNLQVEEKGTAVFTCKTEHPAATVTVRKGLLBELRASKGHPQSQ 240

Qy 2966 EGTLSLTLSALEKADSDTYTCDIGQAQRAQLLVQGRVHHIEDLEDVDVQEGSSATFR 3025  
Db 241 EGTLSLTLSALEKADSDTYTCDIGQAQRAQLLVQGRVHHIEDLEDVDVQEGSSATFR 300

Qy 3026 CRISPANYPFVHFWLDPKTLPHANELNEIDAQPGYHVLTLRQALAKDSGTIYFEAGDQRA 3085  
Db 301 CRISPANYPFVHFWLDPKTLPHANELNEIDAQPGYHVLTLRQALAKDSGTIYFEAGDQRA 360

Qy 3086 SAALRVTEKPSVFSRSLTDATITTEGEDTLVCTSTCDIPMCWTGDKTLRGARCOLSH 3145  
Db 361 SAALRVTEKPSVFSRSLTDATITTEGEDTLVCTSTCDIPMCWTGDKTLRGARCOLSH 420

Qy 3146 EGHRAQLLITGATLQDSGRYKCBAGACSSSIVRVHARPVRFOEALKDLEVLGGGAATLR 3205  
Db 421 EGHRAQLLITGATLQDSGRYKCBAGACSSSIVRVHARPVRFOEALKDLEVLGGGAATLR 480

Qy 3206 CVLSSVAAPVKWCYGNVLPBGDKYSLRQEGAMLELVNRLRPDQSGRYSCSGDQTTSA 3265  
Db 481 CVLSSVAAPVKWCYGNVLPBGDKYSLRQEGAMLELVNRLRPDQSGRYSCSGDQTTSA 540

Qy 3266 TLTVTALPAQFIGKLRNKEATEGATATLRCELSTKAPVVEWRKSETLRDGDYCLRDGA 3325  
Db 541 TLTVTALPAQFIGKLRNKEATEGATATLRCELSTKAPVVEWRKSETLRDGDYCLRDGA 600

Qy 3326 MCELQIRGLAVDAAEYSVCVGBERTSASLTIRMPAHFTIRGRUHOESIGATATLRCEL 3385  
Db 601 MCELQIRGLAVDAAEYSVCVGBERTSASLTIRMPAHFTIRGRUHOESIGATATLRCEL 660

Qy 3386 SKAAPVEWRKGRSLRDGDRHSLRDGAVCELOICGLAVADAGEYSCVCEERTSATLTV 3445  
Db 661 SKAAPVEWRKGRSLRDGDRHSLRDGAVCELOICGLAVADAGEYSCVCEERTSATLTV 720

Qy 3446 KALPAKTEGIRLNEEAVEGATAMLWCLSKVAPVVEWRKGPENLRDGDYCLRDGTRCEL 3505

Db 721 KALPAKFTGLRNEBAVEGATAMLNCELSKVAPVEWRKGPENLRDGRYILRQEGTRCEL 780  
QY 3506 QICGLAMADAGEYLCVCGQERTSATLITRALPARFIEDVKQCEAREGATAVLOCELNSAA 3565  
Db 781 QICGLAMADAGEYLCVCGQERTSATLITRALPARFIEDVKQCEAREGATAVLOCELNSAA 840  
QY 3566 PVWRKGSSTLRDGRYSRLQDGTQKCELIQIRGLAMADTGEYSCVCGQERTSAMLTVRALP 3625  
Db 841 PVWRKGSSTLRDGRYSRLQDGTQKCELIQIRGLAMADTGEYSCVCGQERTSAMLTVRALP 900  
QY 3626 IKFTGLRNEBATEGATVLRCELSKMAPVWPKGHETLRDGRHSRLQDGRARCELOIRG 3685  
Db 901 IKFTGLRNEBATEGATVLRCELSKMAPVWPKGHETLRDGRHSRLQDGRARCELOIRG 960  
QY 3686 LVAEDAGEYLCMGKERTSAMLTVRAMPKFTIEGLRNEEATEGDTATLWCELSKAAAPVEW 3745  
Db 961 LVAEDAGEYLCMGKERTSAMLTVRAMPKFTIEGLRNEEATEGDTATLWCELSKAAAPVEW 1020  
QY 3746 RKGHETLRDGRHSRLQDGRSCELOIRGLAVVDAGEYSCVCGQERTSATLITVRALPARFI 3805  
Db 1021 RKGHETLRDGRHSRLQDGRSCELOIRGLAVVDAGEYSCVCGQERTSATLITVRALPARFI 1080  
QY 3806 EDVKQCEAREGATVLOCELSKAAAPVEWRKGSSETLRGGDRYSRLQDGTQKCELIHGLSVA 3865  
Db 1081 EDVKQCEAREGATVLOCELSKAAAPVEWRKGSSETLRGGDRYSRLQDGTQKCELIHGLSVA 1140  
QY 3866 DTGEYSCVCGQERTSATLITVRAPQVFPREPLQSOABEGSTATLOCELSSEPTATVWWSKG 3925  
Db 1141 DTGEYSCVCGQERTSATLITVRALPARFTQDLKTEASEGATATLOCELSK-VAPVEWKG 1199  
QY 3926 GILQANGRRPRLOQCTAELVLODQREDTGEYTCGSOATSATLITVTAAPVRLREL 3985  
Db 1200 PETLRDGRYSRLQDGTQKCELIHGLSVAADAGEYSCVCGQERTSATLITVRALPARFTEGL 1259  
QY 3986 QHVEVDEGTAHLCELSRAGASVEWRKGSLOLPFCAYQWQDGAALVIRGVEQEDA 4045  
Db 1260 RNEEAMEGATATLOCELSKA-APVEWRKGLEALRDGDKYSRLQDGAVALCELIHGLAMADN 1318  
QY 4046 GDYTCDTGHTOSMASLSVRVPRPKFTLRLOLEQETGDIARLCCOLSDAESGAVVOWLKE 4105  
Db 1319 GYSCVCGQERTSATLITVRALPARFIEDMRNQKATEGATVTLQCKLRKA---APVEWRKG 1375  
QY 4106 GVLEHAGPKYEMRSQCATRELLIHQLEAKDTGEYACVVGQKTAASLRVTEPEVTIVRGL 4165  
Db 1376 PNTLRDGRYSRLQDGTQKCELIHGLSVAADAGEYSCVCGQERTSATLITVRALPARFIEDV 1435  
QY 4166 VDAEVTADEDVFEFCEVSRAGATGVQWCLQGLPLQSNVTEVAVRGRHITLRLKGVTP 4225  
Db 1436 RNHEATEGATVLOCELSKAAAP--VEWRKSETLRDGRYSRLQDGTQKCELIHGLAVE 1492  
QY 4226 DAGTVSFHLGNHASSAQITVRAPVITILEPLQDVQLESEGQDASFOCKLSRASGOEARWAL 4285  
Db 1493 DTGEYLCVCGQERTSATLITVRALPARFIDNNTQCEAREGATATLHCELSKVAPVE--WRK 1550  
QY 4286 GGVPLQANEMNDITVEQGTLLHLTLHKVTELDAGTVSFHVGTG 4328  
Db 1551 GPETLRDGRHSRLQEN-----RLNPGGCGCSELGSC 1582

Search completed: September 13, 2004, 12:12:13  
Job time : 541 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 14:16:08 ; Search time 1570.02 Seconds  
(without alignments)  
10628.572 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_16862\_17246

Perfect score: 385

Sequence: 1 ccagagcgagatcttgaca.....acctgcagcactggagcgc 385

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 347072 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*

2: gb\_hlg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vi:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vi:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htg\_mus:\*

34: em\_htg\_pln:\*

35: em\_htg\_rod:\*

36: em\_htg\_man:\*

37: em\_htg\_vrt:\*

38: em\_sy:\*

39: em\_htgo\_hum:\*

40: em\_htgo\_mus:\*

41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	385	100.0	4901	9	HS803665	AL832357 Homo sapi
2	385	100.0	7928	6	AX039412	AX039412 Sequence
3	385	100.0	20435	9	HSJ2535	AJ002635 Homo sapi
4	377	97.9	628	6	AX308286	AX308286 Sequence
5	274.8	71.4	4171	10	BC060226	BC060226 Mus muscu
6	204.8	53.2	23123	9	HSJ14908	AJ314908 Homo sapi
7	204.8	53.2	135864	9	AL353593	AL353593 Human DNA
8	180.8	47.0	164766	2	AC026657	AC026657 Homo sapi
9	171.4	44.5	250087	2	AC096931	AC096931 Rattus no
10	171.4	44.5	260998	2	AC099089	AC099089 Rattus no
11	163.4	42.4	211829	10	AL645854	AL645854 Mouse DNA
12	93.8	24.4	174612	2	AC023889	AC023889 Homo sapi
13	85	22.1	98613	2	AC139630	AC139630 Takifugu
14	48	12.5	349981	1	BX572602	BX572602 Rhodopsu
15	47.4	12.3	300425	1	AP005022	AP005022 Streptomy
16	43.2	11.2	296500	1	SC0939128	AL939128 Streptomy
17	42.6	11.1	4180	9	HUMB94	M92357 Homo sapien
18	42.2	11.0	1553	8	TAB276509	AJ276509 Tricium
19	42.2	11.0	191996	9	AC092275	AC092275 Homo sapi
20	41.4	10.8	1803	6	BD276572	BD276572 EXTRACELL
21	41.4	10.6	169824	2	AX048206	AX048206 Sequence
22	41	10.6	158462	9	AL592301	AL592301 Human DNA
23	40.6	10.5	1078	1	PATPPIG	X51868 Pseudomonas
24	40.4	10.5	19372	2	AE004443	AE004443 Pseudomon
25	40.4	10.5	19372	2	AE004443	AE004443 Pseudomon
26	40.2	10.4	145577	2	AC145762	AC145762 Sus scrof
27	40.2	10.4	194809	4	AC091506	AC091506 Sus scrof
28	40.2	10.4	292100	1	SC0939121	AL939121 Streptomy
29	40.2	10.4	348525	1	BX640428	BX640428 Bordetell
30	40.2	10.4	348642	1	BX640446	BX640446 Bordetell
31	40.2	10.4	348997	1	BX640427	BX640427 Bordetell
32	40	10.4	299800	1	AP005040	AP005040 Streptomy
33	39.8	10.3	125020	9	AF429315	AF429315 Homo sapi
34	39.6	10.3	9199	1	AY256972	AY256972 Pseudomon
35	39.6	10.3	151159	2	AC145010	AC145010 Sus scrof
36	39.6	10.3	170416	2	AC145011	AC145011 Sus scrof
37	39.4	10.2	1341	9	BC006223	BC006223 Homo sapi
38	39.4	10.2	2040	6	AX747647	AX747647 Sequence
39	39.4	10.2	2040	9	AK092842	AK092842 Homo sapi
40	39.4	10.2	123580	1	AF263912	AF263912 Streptomy
41	39.4	10.2	125401	6	AX211739	AX211739 Sequence
42	39.4	10.2	182662	9	AC023830	AC023830 Homo sapi
43	39.4	10.2	207961	2	AC013282	AC013282 Homo sapi
44	39.4	10.2	222643	2	AC146802	AC146802 Silurana
45	38.8	10.1	3001	5	LFL308113	AJ308113 Lampetra

# ALIGNMENTS

RESULT 1  
LOCUS HS803665 4901 bp mRNA linear PRI 13-MAY-2003  
DEFINITION Homo sapiens mRNA; cDNA DKFPZ451F056 (from clone DKFPZ451F056).  
ACCESSION AL832357  
VERSION AL832357.1 GI:21732919  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 4901)  
AUTHORS Koehrer,K., Beyer,A., Wewes,H.W., Weil,B., Amid,C., Osanger,A.,  
Fobo,G., Han,M. and Wiemann,S.  
TITLE Direct Submission

JOURNAL Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764  
Neuberberg, GERMANY  
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@kfz-heidelberg.de;  
sequenced by BMFZ (Biomedical Research Center at the  
Heinrich-Heine-University, Dueseldorf/Germany) within the cDNA  
sequencing consortium of the German Genome Project. This clone  
(DKFZ451F056) is available at the RZPD in Berlin. Please contact  
the RZPD; Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clones@rzd.de Further  
information about the clone and the sequencing project is available  
at http://mips.gsf.de/proj/cDNA/.

FEATURES  
Location/Qualifiers  
1..4901  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="RZPD:DKFZp451F056"  
/db\_xref="taxon:9606"  
/clone="DKFZp451F056"  
/tissue\_type="human skeletal muscle"  
/clone\_lib="451 (synonym: hlcc1). Vector pSport1; host  
DH10B; sites NotI + SalI"  
/db\_stages="adult"  
/db\_xref="GenBank:U000000000"  
polyA\_signal 4828..4833  
polyA\_site 4847

ORIGIN  
Query Match 100.0%; Score 385; DB 9; Length 4901;  
Best Local Similarity 100.0%; Pred. No. 2e-71;  
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CCGAGGCGAGATCTTTGACATCTACGTGCTACCGGCTGACTACCTGCGCCCTAGGGCTGA 60  
Db 1281 CCGAGGCGAGATCTTTGACATCTACGTGCTACCGGCTGACTACCTGCGCCCTAGGGCTGA 1340  
Qy 61 GCAGATGCCATCAGCTGCGGAGGAGCCAGTATGTGGAGTCTTGATGTCACGCCACCC 120  
Db 1341 GCAGATGCCATCAGCTGCGGAGGAGCCAGTATGTGGAGTCTTGATGTCACGCCACCC 1400  
Qy 121 ACTCGCTGGCTGTCGCCAGCAGGAGGCTCAAGCTGTCACCTGAGTGGGGGCGCTGAGC 180  
Db 1401 ACTCGCTGGCTGTCGCCAGCAGGAGGCTCAAGCTGTCACCTGAGTGGGGGCGCTGAGC 1460  
Qy 181 GTCCACGAGCTTACCTGACAGGAGGCTCAAGCTGTCACCTGAGTGGGGGCGCTGAGC 240  
Db 1461 GTCCACGAGCTTACCTGACAGGAGGCTCAAGCTGTCACCTGAGTGGGGGCGCTGAGC 1520  
Qy 241 CCCTGAGTTCCTGGGAGGCTGTGCTGAAGACGAATCAAGGCAAGGCTGAGCTCTGT 300  
Db 1521 CCCTGAGTTCCTGGGAGGCTGTGCTGAAGACGAATCAAGGCAAGGCTGAGCTCTGT 1580  
Qy 301 GATCCAGAGCTGCTGAGTTCGAGGAGGCTGCTGAGGAGGCTGCTGAGGAGGCTGCTGAGG 360  
Db 1581 GATCCAGAGCTGCTGAGTTCGAGGAGGCTGCTGAGGAGGCTGCTGAGGAGGCTGCTGAGG 1640  
Qy 361 CCACACTGCGAGCAGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGG 385  
Db 1641 CCACACTGCGAGCAGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGG 1665

RESULT 2  
AX039412  
LOCUS AX039412  
DEFINITION Sequence 5 from Patent WO0063381.  
ACCESSION AX039412  
VERSION AX039412.1 GI:11229480  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Zeng, W., Stanton, L. and Kong, H.

TITLE Mammalian protein with putative function in signal transduction  
JOURNAL Patent: WO 0063381-A 5 26-OCT-2000;  
SCIOS INC. (US)  
FEATURES  
Location/Qualifiers  
1..7928  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
60..7850  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC16626.1"  
/db\_xref="GI:11229481"  
/translat:cn="MLERFPTPKVKGSGITFSVKVGRDPTVTHLREAEAGVLMI  
GPDTPGYTVASSAQHSLVLLDVGROHGYTCTAGNAQALCSALHVSGLPKVEE  
QEKVKEALISTFLQGTTCATSAOGLFETADLGQKKEBPAALAKALHSLAEVVE  
EEFLQKLTSTITKWSAKITQAKLOVPGSDSDSKTSPSPHSGRSPSSIOESST  
ESDGDARGEIIVYVITDILPLGASQDAITREGQIVSVDAAPLRLVYTKTK  
SSPQRCQWSPAYDLRLKLSPEWGAEEAEPFEVGEVDEYKARUSSVQELLSSEQ  
AFVEELQASHHLQHLRCPHPIAVAGQKAVIFRNRDIDGRFHSFVQELQCCOTD  
PPLQHYLEQPAVORVYQALLKELIRNARNONCALLBOAVAVSALPORAENKLHV  
SLMENYPTLEALGEP:ROGHFTWEGACGARMKGNHNVFLFRNHLVICKPRDS  
RTDVSVYFRNMKLSIDANDQVEGDRAFEVWQEREDSVRYKYLLOARAIKSNW  
KEICIQOURLALPVRPPDFEELADCTAELGTVTKLACRVGTGTPVPIWYKDGKAV  
QVDPHHILIEDPDGSCALILDSTVDSDGYMCAASAGNCSTLGLKILVQVPPRVN  
KVRASFEVEGEADAQFTICEAPYQIRWYDKGALLTGNKFTLSEPPSGLLIVIR  
AASXEDLGLYCELVNRLGASASAEIRI QSPMLQAEQCHREQLVAIVEDTLEARD  
QEVTSVLRLKPKAPGPGSTDLTGPOCPRGAPALQETGSPVVTGTEAPVPRV  
POILLHSGPQEPPEAIARAQEWTVPIRMEGAWEAGTGLLMDVSHVVRVETQTY  
TYQADITHRTASPMQVITIEDVQAQGTGTAQFEALLEGDPQSPVWYKQSVLVDSTR  
LSQOQEGTYSLVLRHVASKDAGVYTCIAQNTGQVLCALVLLVGLDNEPDSKOSH  
RKJHSFYVEKEIGRQVGVFVVRVQKSHNKLICAAKIFIPRSRTAQAIRQEDILAA  
LSHPVLTLGDDQFETRTLLILELSEKSELRLYRKGVVTEAEVKYIQFLEVEHL  
YQMPVSEASDIWAMGVIYSULTCSSPFAGSDRATLLNLVGRVSWSPMAHISE  
DAKDFIKATLQAPAESAOCLSHPWFKSMPEAEAFINTKOLKFLRLARWORS  
LMSYKSLVWRSIPELLRGPDSPLGVABHLCBDTGGSSSSSSSSSNELEAPARAKS  
LPSPFVPLHPRGFRFSASLPEAEASERSTEAPAPAPSPAGNPPAAAGCVPR  
HSVIRSLFYQAGSEPHGALAPGRHRRHLLKGGYIAGALGRLPREPLMEHRLV  
EEAEARQATLLAKAPSFETALPASPHTLAPGSHSLHSDSPSPSPSPSACCEA  
QRLSPASGAPADRMHGPQSKPLSTGGHPTAQPERSPSPSPSPSPSPSPSPSPSP  
SAPEGCSPPAVAPCPGSPGSCKEAPLVPSSPPLGQPPAPAPAPAPAPAPAP  
GPGDISLPGRKPGKSPGSAOASOSSVSSVLRVSSOVGTGTPSLDAEGMTQAE  
DLSTSTLQRPQATWYKXESLGGRGYAGVAGYGTFAFGDAGMGLGOGPWARI  
WAVSSEBEEQEAASQSEBQEAESPLQVRSARVPYEVGRAPTRSPSPSPSP  
DIGQVSLVQIRLSDGDAEDTSLDISVDPAFLNLSLDYDKILFFEFNIFKPKK  
SAQPEPSMAEEAEFPPTWPMGELPHAGLEITEESVDLALABAAVGRKKK  
WSPSRSLFFPGRHPLDPLPAELGRVRYKASVEHSIRILKGRPELEKEGPPRKP  
GLASFRLSKLMDRPTLRELSDETGLQSVTLACQVSAQAPAAQAKXGDALE  
SSSVLLSATLKNFOLLTIIVVAEDLGVYTCVSNALGTVTVTGVLURKABRSPPC  
PDIGEVADGVLYWKPVEYGPVTYVQCSLEGGSWTTLASDIFDCQYTLTSLSRGG  
SPRHLVILELCSGPELLPLCLAEASVSESEVDYLMQMSATQYLNHQLHLDLRS  
ENMIITEYNLLKVDLGNASQELVKDFKDYLETWAPLELQGGQVAPQDIWA  
IGVTAFIMLGAEPVSEGDRLQRLKGLVRLSCYAGLSGGAFLRSTACQW  
GRPCASSCLQCPWLITBEGSPACRPAPVFTFTARLVRVFRNRKRLLYKRNLAQVR

ORIGIN  
Query Match 100.0%; Score 385; DB 6; Length 7928;  
Best Local Similarity 100.0%; Pred. No. 1.9e-71;  
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CCGAGGCGAGATCTTACATCTACGTGGTCCACCGCTACCTACCTGCCCTAGGGCTGA 60  
Db 734 CCGAGGCGAGATCTTACATCTACGTGGTCCACCGCTACCTACCTGCCCTAGGGCTGA 793  
Qy 61 GCAGGATCCCATCGCTGCGGGAAGCCAGTATGTGGAGGTCTGTGATGAGCCACCC 120  
Db 794 GCAGGATCCCATCGCTGCGGGAAGCCAGTATGTGGAGGTCTGTGATGAGCCACCC 853

QY 121 ACTGGCTGGCTTGTCCGACCAAGCCACCAAGTCCAGGCCCTCAGCGAGGCTGGGT 180  
 Db 854 ACTGGCTGGCTTGTCCGACCAAGCCACCAAGTCCAGGCCCTCAGCGAGGCTGGGT 913  
 QY 181 GTCACCAAGCTTACTGACAGGAGGCTCAAGCTGTCACTGAGTGGGGGCGCGTGAAGC 240  
 Db 914 GTCACCAAGCTTACTGACAGGAGGCTCAAGCTGTCACTGAGTGGGGGCGCGTGAAGC 973  
 QY 241 CCCTGAGTTCCTCGGAGGCTGTCTGAAGACGAATACAAAGCAAGGCTGAGCTCTGT 300  
 Db 974 CCCTGAGTTCCTCGGAGGCTGTCTGAAGACGAATACAAAGCAAGGCTGAGCTCTGT 1033  
 QY 301 GATCCAGAGCTGCTGAGTTCGAGCAGGCTTCGTCGAGAGAGCTGCAGTTCCTGACAG 360  
 Db 1034 GATCCAGAGCTGCTGAGTTCGAGCAGGCTTCGTCGAGAGAGCTGCAGTTCCTGACAG 1093  
 QY 361 CCACCACTGCAGCAGCTGGAGCGC 385  
 Db 1094 CCACCACTGCAGCAGCTGGAGCGC 1118

RESULT 3  
 HSAJ2535 20435 bp mRNA linear PRI 14-SEP-2001  
 LOCUS Homo sapiens mRNA for obscurin (OBSN gene).  
 DEFINITION  
 ACCESSION AJ002535  
 VERSION AJ002535.1 GI:15026973  
 KEYWORDS OBSN gene; obscurin.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1  
 AUTHORS Young, P., Ehler, E. and Gautel, M.  
 TITLE Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor  
 protein involved in sarcomere assembly  
 JOURNAL J. Cell Biol. 154 (1), 123-136 (2001)  
 MEDLINE 21342081  
 PUBMED 11448995  
 REFERENCE 2 (bases 1 to 20435)  
 AUTHORS Gautel, M.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-OCT-1997) Gautel M.S., Structural Biology Division,  
 European Molecular Biology Laboratory, Meierhofstr. 1, Heidelberg,  
 69117, GERMANY  
 REMARK Revised by author 20-JUL-2001  
 FEATURES  
 Location/Qualifiers  
 1..20435  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="1q42.1"  
 /cell\_type="cardiomyocyte"  
 /tissue\_type="cardiac"  
 /dev\_stage="adult"  
 1..20435  
 /gene="OBSN"  
 /gene="OBSN"  
 /function="titin binding"  
 /codon\_start=1  
 /evidence="experimental"  
 /product="obscurin"  
 /protein\_id="CAC44768.1"  
 /db\_xref="GI:15026974"  
 /db\_xref="GOA:Q96AA2"  
 /translation="MDQPOFGSAPREFLTPKAFVSVGKDATLSQIVGNPTQVSWSE  
 KDQQVAFGRFLRGLDGLYRLTLDLALGSGGVVCPARNAIGAPFAVGLVDAAE  
 ACAEQAPFLLRPTIRVREGSEATFCRVGSGPRPVSWSKDGRLGEPDGRVY  
 EEUGASALRIRARPRDGGTYVRAENPLGAASAAALVVDSDAADTASRFGTSTA

source

gene

CDS

LLAHLORREARMRAEGAPASPESTGTRTCTVTEGKHARLSGVVTEGPKPVTWKKGDL  
 LVTEGRHVVVDAQENFVLKILFCKQSDRLGYLTCTASNLVGTQYSSVLVVRPAVP  
 PKRLQDLVREKESATFELCEVPQPSSTEAAMFKETRLWASAKYIEGEGTERRLTVR  
 NVSADDVAVICETPEGSKTVAELAVQNLRLKPRKTAVRGDTAMFCVELAVPVG  
 VHLNRQBEVWAGGRVIAISAGTRHTLTIISQCLLEDVQVAFPMAGDGTSTFCVSP  
 RPPDPVDPVVKARMESSVLISWSPPHGPERPVTIDIGLYEKKKGLTYTWRICHEA  
 EWVATPELTADVAAEGNFQVRSALNSFGQSPYLFPGTVELAPLAVRPLKAVQA  
 VEGGVTVSDLVTVASAGFWLDQALKASSVYETHCDRTTRHTLTIRVPSALHQAOL  
 KVFANGIESSIRMEVRAAPLGTANKPPAAAREVILARHEAOLLAELSDQAAATML  
 KDGRTLSGPKKVEVQASAGRRVLLVRDVARDDAGLYECVSRGGRIAYQLSQGLARFL  
 KDMAGSCDVAAGVAGPQFECETSEAHVHVWYKDMBELHSGERFLDVGTRHLV  
 AATVTRQDEGTYSCRVDGDFRFRVSEPKVVFVFAKEQLARRKQAAGASATLSCEV  
 ADQATEWTYKDGKLLSSSKVCEATCTRLVVQAGADAGEYSCAAGQRLSFH  
 LDKEPQVFAKDVAHSEVQAEAGANATLSCEVAQAEVWYKDGKLLSSKLVHV  
 EAKGRRLVVQAGKTGDAGYSCEARGORVSRFLHTEPKMFAKESVNEVQAEABA  
 GASMLSCSEVQAAQTEVTYKDGKLLSSSKVMEVKCTRLVLPQAKGADAGEYSC  
 EAGGORVSHLHITPEKGVFAKEQSVHNEVQAEAGTTAMLSCEVQAPOTETVWKDGK  
 KLSSSKVMWVKGTCTRLVVQAGADAGEYSCAAGQRLSFHITPEKAVFAKEQ  
 LVHNEVTEAGASATLSCEVQAAQTEVTYKDGKLLSSSKVIRIQAAGQRLVVQQA  
 GOADAGEYTCAGGQRLSFHLDVSEPKAVFAKEQLARRKQAAGASATLSCEVQAAQ  
 TEVTYKDGKLLSSSKVMEVAVGCTRLVVQAGADAGEYSCAAGQRLSFSDVA  
 EPKVFAKEQPVHREVQAGASTTSCVQAAQTEVTYKDGKLLSSSKVMEVAVG  
 CTRLLVQAGADAGEYSCAAGQRLSFHLDVSEPKAVFAKEQSVHNEVQAEAGTSA  
 TLSCEVQAAQTEVTYKDGKLLSSSKVMEVAVGCTRLVVQAGADAGEYSCAAGT  
 QRLSFHLDVSEPKAVFAKEQPAHREVQAEAGASATLSCEVQAAQTEVTYKDGKLLS  
 SKVMEVAVGCTRLVVQAGADAGEYSCAAGQRLSFHLDVSEPKAVFAKEQSVHNEV  
 PLVKEHEDILITATLTPSAATVTLWKGVEIRSRKHETASQDHTLTVHQAQVL  
 DSALYSCRVGAEGQDFPVQVEEVAKFCRLLEPVCGLGTTVLACELSPACAEVWVR  
 CGNTOPVTVKRGFQVAVGPRVSLTVGLRAEDAGEYVCEVCEDDHTSAOLTVSVPRVVK  
 FMSGISTVVEBEGEATPCQVSPSDVAVFRGALLQPSKFAISGASGASHLITIS  
 DLVLEDAQITVEAGSASSAALVRPAVPLFKKLEPQTVVEEISSVTEVELTRPWP  
 ELRTRNATALAPGNVEITHAEGARHLVHLNVGFADRGFCGCTPDDTKAKUTVM  
 RQVRLVQLQVAREQGTATMEVHSHADVDGNTDRGLRFQOQGTCHLVAQGMHT  
 LTLGLRDESDGLAVFAEGVHTSARLVVTELPVFSRPLQDVVTEKVTLECLS  
 RPNVDVRLKDGELVRAKGTMAIAAQGACRSLLTYRCEFAQGVQVQVCAHDAQSSAS  
 KVQGRYTLIVRVLAEDAGIIOFVAENESRAOLRVKELPVLVPLRDLKAMEKHR  
 GLVCEQVRSASQVRWFKGSELOQPKVVELVSDGLVKLLISDVHREDTDTLQAG  
 DVKTSAQFFVEEQSITIVRGQDVTVMPEZAPAMPECETSIISVRPPKMLLAKTVLQAG  
 GNVLEGGTVHRLMLRKTCTMTGTGTVKGRSSARLVSDIPVVTLPLEPPTG  
 RELQSVLSCDFRPAPKAVONYKDDTPLSPSEKFMSLGQMGLRLRLPMDAGVY  
 RCQAGSAHSTETVAREVTVTGPLQDAEATEGWASFCSELSHEDSEVWSLNGMP  
 LYNDSEFHEISHKRRRLTVLXSIORADAGIVRASLSKVSFSTARLEVRVVPVVKALD  
 DLSAEROTLALOCEVSPDEAHVWRKGVOLGSPDKYDLFLHATAGTGLVHDVSDPED  
 AGLTYCHVGSSETRARVRVHDHVGITKRLATMEVLEGESCFEVLISHESASBPAMW  
 TVGKTVSSSRFQATQGRKXILVREAAFSADAGEVSVRGLTSKASLVRBRPAA  
 I1KPLEDQWVAPGEDVELRCELSRAGTPVHLKDRKIRKSKYDVVCEGVMAMLVIR  
 GASLKADGEYTCVEASKSTASLHVEKANCFTBELTNLQVEEKGTAFTCKTHPPA  
 TVWRKGLLELRASGKHQPSQGLTLRLTISALEKADSDTYTCDIGQASRAQLLVQ  
 RVVHIELEDVDVQEGSSATFCRISIPANYEPVWFJDKTPLHNEIDAOQVGY  
 HVLTROLALDGTIYFEAGDORASALRVTEKPSVRELTDTATITEGREDTLVCE  
 TSTCDIPMCWTKDGTILRGSAQCQLSHEGHAQLITATLQDSGRYKCEAGGACSS  
 IYVRHARVRFQEAALKULEVLEGAATLRCVLSVAPVKWCIYGNVLRPGDKYSLRQ  
 EGAMLELVNRLPQDSGRYSCSFGDQTTSATLVTALPAQFIKLRNKEATEGAT  
 LRCELSKAPVWRKGSSETLDGRYCLROGDAMCELIQIRGLVDAEYSCVCEER  
 TSASLTIRPMPAHPFIRGLHOESIEGATLRLCSKAAPVWRKGRSLDGRHSL  
 RODGAVCLOICGLAVADAGEYSCVCEERTSATLTVKALPAKTEGLRNEEAVGAT  
 AMLCELSKAPVWRKGPENLRDGRVILRQEGTCELOICGLAMADAGEVLCVCGO  
 BETSATLIRALPARFIEDVQVQAEAGATVLOELNSAPVWRKGSSETLRGDY  
 SLRQGTKEICQLRGLVADAGEYSCVCGQRTSAMLTVRALPFIKTEGLRNEATEG  
 ATAVLRCELSKAPVWRKGSSETLRDGRHSLRQDGRARCELIQIRGLVADAGEYLCMC  
 CKERTSAMLTVRMPKSFIEGLRNEATEGATLWCELSKAAPVWRKGRSLDGRHSL  
 RSLRQDGRCELOIRGLVADAGEYSCVCEERTSATLTVRALPARFIEDVQVQAE  
 EGATAVLQCELSKAAPVWRKGSSETLRDGRYSLRQDGRCELOIRGLVADAGEYSC  
 VCGERTSATLTVRAPQVPRPQLSQAEGSTATLQCELSPTAVVWSKGLQLQ  
 ANGREPRLOCTAELVLOLQREDTGEYTCGSOATSATLTVAPVPRFLRLQHQ  
 EVDEGTAHLCCELSRAGASVWRKGSQVPCAKYQMVQDGAELLVRGVEDEAG  
 DYTCDTGHQTSMSLSVRVPRPKTKRLQSEGTGDIARLCCQLSQAEGVQWMLK  
 EGVELHAGPKYEMRSQAGATRELLIHOEAXDTGEYAGTGGTAAASLRVTEPVTIV  
 RGLVDAEYTCVCEVSRAGATVQVCLQPLQSNSETEVAVDGRIRHLRLK  
 RGLVDAEYTCVCEVSRAGATVQVCLQPLQSNSETEVAVDGRIRHLRLK  
 QVTPEDAGTVSHLNGSAAQTVRAPEVTLIPLOQVDSQEGDASFOCRILGRASG  
 CEASVWGLVPLQANMNDLIVEOGTILHLTHKVTLEDAGTVSHFGVTSSEALQV  
 TAKNTVWGLVPLNEALSGGALFECQLSQEVAATHLLDDEPVRTSENARVVFENG  
 LRHLLLLKNLRPQDSCTVFIAGDNVTSFALTVRGRWRLDEILEPUNAAVRAQRFIT







TRRSADSCQSSFTTELSTQTQVNFVGTGTTLHICPRDRGDGAAP

```

exon      15192..15303
/gene="OBSCN"
/number=85
intron    15304..18393
/gene="OBSCN"
/number=85
exon      18394..18436
/gene="OBSCN"
/number=86
intron    18437..21834
/gene="OBSCN"
/number=86
exon      21835..23123
/gene="OBSCN"
/number=87

ORIGIN
Query Match      53.2%; Score 204.8; DB 9; Length 23123;
Best Local Similarity 99.0%; Pred. No. 1.8e-33;
Matches 206; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 AGATCTTTGACATCTAGCTGCTACCGCTGACTACCTGCTCCCTAGGCGCTGAGCAGGATG 68
Db 210 AGATCTTTGACATCTAGCTGCTACCGCTGACTACCTGCTCCCTAGGCGCTGAGCAGGATG 269
QY 69 CCATCAGCTCGGGAGGCCAGTATGTGGAGTCTCTGGATGCGAGCCGCCACTGGCT 128
Db 270 CCATCAGCTCGGGAGGCCAGTATGTGGAGTCTCTGGATGCGAGCCGCCACTGGCT 329
QY 129 GCTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGGCGGCTGGGTGTGCACAG 188
Db 330 GCTTGTCCGACCAAGCCACCAAGTCCAGCCCTCAGCGGCGGCTGGGTGTGCACAG 389
QY 189 CCTACCTGGACAGGAGGCTCAAGCTGTC 216
Db 390 CCTACCTGGACAGGAGGCTCAAGGTATC 417

RESULT 7
AL353593 Human DNA sequence from clone RP5-1139B12 on chromosome 1q42.1-43,
LOCUS complete sequence.
DEFINITION AL353593
ACCESSION AL353593.33 GI:18673899
VERSION HTG
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 135964)
AUTHORS Almeida J.
TITLE Direct Submission
JOURNAL Submitted (13-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
On Feb 14, 2002 this sequence version replaced gi:17977879.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
```









```

SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL      1 Smith, M.
TITLE        Direct Submission
COMMENT      Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
             Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
             humquary@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
             On Apr 7, 2002 this sequence version replaced gi:119699553.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence was finished as follows unless otherwise noted: all
             regions were either double-stranded or sequenced with an alternate
             chemistry or covered by high quality data (i.e., phred quality >=
             30); an attempt was made to resolve all sequencing problems, such
             as compressions and repeats; all regions were covered by at least
             one plasmid subclone or more than one M13 subclone; and the
             assembly was confirmed by restriction digest. The following
             abbreviations are used to associate primary accession numbers given
             in the feature table with their source databases: Em., EMBL; Sw.,
             SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
             database can be found at
             http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-192P17 is
             from the RP23-23 Mouse PAC Library
             constructed by the group of Pieter de Jong.
             For further details see http://www.chori.org/bacpac/home.htm
             VECTOR: pBac3.6.

FEATURES             Location/Qualifiers
     source           1..211829
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clones="RP23-192P17"
                     /clone_lib="RPCI-23"

ORIGIN
Query Match      42.4%; Score 163.4; DB 10; Length 211829;
Best Local Similarity 87.3%; Pred. No. 6.7e-25;
Matches 179; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 9 AGATCTTTGACATCTACCTGGTCAACCGTGACTACCTGCCCTAGGGGCTGAGCAGGATG 68
    |||||
Db 15366 AGATCTTTGACATCTATGTGTGTCACAGTGACTATCTGCCACTGGGAGCTGAGCAGGATG 15307

Qy 69 CCATCAGCTCTCGGAAGGCCAGTATGTGGAGTCTCTGGATGTCAGCCACCCACTGCGCT 128
    |||||
Db 15306 CCATCATTCTGAGAGAAGGCCAGTATGTGGAGTCTCTTGACTCTGCCCATCCCTCGGCT 15247

Qy 129 GCGTTGTCGACCAACCAAGCCAGCCAGTCCAGCCCTCAGCGGAGGGTGGTGTCCACGAG 188
    |||||
Db 15246 GCGTTGTCGAGACCAAGCCACCAATCCAGTCTCTCCAGCGAGGGTGGTGTTCACCTG 15187

Qy 189 CCTACCTGGACAGGAGGCTCAAGCT 213
    |||||
Db 15186 CCTACCTGGATAGAGGCTCAAGT 15162

RESULT 12
AC023889/c
LOCUS      AC023889.3 174612 bp DNA linear HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-661B12, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION AC023889
VERSION   AC023889.3 GI:8969253
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE    Homo sapiens (human)

```

---

```

ORGANISM    Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
JOURNAL      1 (bases 1 to 174612)
TITLE        The sequence of Homo sapiens clone
COMMENT      Unpublished
             2 (bases 1 to 174612)
             Waterston,R.H.
             Direct Submission
             Submitted (18-FEB-2000) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
             On Jul 7, 2000 this sequence version replaced gi:8748947.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0661B12
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing method: plasmid; 0%
Chemistry: Dye-terminator ET; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 163945 bases at least Q40
Consensus quality: 167601 bases at least Q30
Consensus quality: 169687 bases at least Q20
Insert size: 148000; agarose-1p
Quality coverage: 5.67 in Q20 bases; sum-of-contigs
Quality coverage: 5.55 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2478: contig of 2478 bp in length
* 2479 2578: gap of unknown length
* 2579 6948: contig of 4370 bp in length
* 6949 7048: gap of unknown length
* 7049 9373: contig of 2325 bp in length
* 9374 9473: gap of unknown length
* 9474 13303: contig of 3830 bp in length
* 13304 13403: gap of unknown length
* 13404 18270: contig of 4867 bp in length
* 18271 18370: gap of unknown length
* 18371 25589: contig of 7219 bp in length
* 25590 25689: gap of unknown length
* 25690 34815: contig of 9126 bp in length
* 34816 45742: contig of 10827 bp in length
* 45743 45843: gap of unknown length
* 45843 62987: contig of 17145 bp in length
* 62988 63088: gap of unknown length
* 63088 87584: contig of 24497 bp in length
* 87585 114765: contig of 27081 bp in length
* 114766 114865: gap of unknown length
* 114866 142334: contig of 27469 bp in length
* 142335 142435: gap of unknown length
* 142435 174612: contig of 32178 bp in length.
             Location/Qualifiers
             1..174612
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"

```

```

/misc_feature 1 /chromosome="1"
/misc_feature 1 /clone="RP11-661B12"
/misc_feature 1 /notes="assembly_name:Contig4"
/misc_feature 2579 /notes="assembly_name:Contig5"
/misc_feature 7049 /notes="assembly_name:Contig6"
/misc_feature 9474 /notes="assembly_name:Contig7"
/misc_feature 13404 /notes="assembly_name:Contig8"
/misc_feature 18371 /notes="assembly_name:Contig9"
/misc_feature 25690 /notes="assembly_name:Contig10"
/misc_feature 34916 /notes="assembly_name:Contig11"
/misc_feature 45843 /notes="assembly_name:Contig12"
/misc_feature 63088 /notes="assembly_name:Contig13"
/misc_feature 87685 /notes="assembly_name:Contig14"
/misc_feature 114866 /notes="assembly_name:Contig15"
/misc_feature 142435 /notes="assembly_name:Contig16"

Query Match 24.4% Score 93.8; DB 2; Length 174612;
Best Local Similarity 97.9%; Pred. No. 3.5e-10;
Matches 95; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 289 GCTGAGCTCTGTGATCCAGAGCTGTGAGTCTTGACGAGCGCTTCTGTGAGGAGTGCA 348
Db 86838 GCCAGCTCTGTGATCCAGAGCTGTGAGTCTTGACGAGCGCTTCTGTGAGGAGTGCA 86779

Qy 349 GTTCTCGACGACCCACCTGACGACCTGAGCGC 385
Db 86778 GTTCTCGACGACCCACCTGACGACCTGAGCGC 86742

RESULT 13
AC139630 LOCUS AC139630.2 GI:28882132
DEFINITION Takifugu rubripes clone 221D8, WORKING DRAFT SEQUENCE, 6 ordered
pieces.
ACCESSION AC139630
VERSION HTG; HTGS PHASE2; HTGS DRAFT.
KEYWORDS Takifugu rubripes (Fugu rubripes)
SOURCE Takifugu rubripes
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Takifugu.
REFERENCE 1 (bases 1 to 98613)
AUTHORS Akhter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakeley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Carriaga, K., Coleman, B., Engle, J., Granite, S., Guan, X., Gupta, J.,
Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E.,
Laric, P., Lee-Hin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B.,
Marquilies, E.H., Masello, C., Maskeri, B., McDowell, J.,
Paguigan, C., Pearson, R., Portnoy, M.E., Prasad, A.,
Reddix-Dugue, N., Schandler, K., Schueler, M.G., Sison, C.,
Stantripp, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Vogt, J.L.,
Wetherby, K.D., Wiggins, L., Young, A. and Green, E.D.
TITLE NISC Comparative Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 98613)

```

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Green, E.D.  
Direct Submission  
Submitted (07-FEB-2003) NIH Intramural Sequencing Center, 8717  
Grovetmont Circle, Gaithersburg, MD 20877, USA  
3 (bases 1 to 98613)  
Green, E.D.  
Direct Submission  
Submitted (08-MAR-2003) NIH Intramural Sequencing Center, 8717  
Grovetmont Circle, Gaithersburg, MD 20877, USA  
On Mar 8, 2003 this sequence version replaced gi:28269380.  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov  
Contact: nisc.zoo@nih.gov  
----- Project Information  
Center project name: egw  
Center clone name: 221D08

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 97841 bases at least Q40  
Consensus quality: 98004 bases at least Q30  
Insert size: 85000; agarose-fp  
Insert size: 98113; sum-of-contigs  
Quality coverage: 11.51x in Q20 bases; agarose-fp  
Quality coverage: 9.97x in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 6 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1 28467: contig of 28467 bp in length  
28468 28567: gap of unknown length  
28568 31135: contig of 2568 bp in length  
31136 31235: gap of unknown length  
31236 41487: contig of 10252 bp in length  
41488 41587: contig of 32778 bp in length  
41588 74365: contig of 32778 bp in length  
74366 74465: gap of unknown length  
74466 95641: contig of 22076 bp in length  
95642 96641: gap of unknown length  
96642 96613: contig of 1972 bp in length.

Location/Qualifiers  
1..98613  
/organism="Takifugu rubripes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:31033"  
/clone="221D8"  
/clone\_lib="Incyte Genomics"  
1..28467  
/note="assembly\_fragment  
clone\_end:T7

misc\_feature

source

FEATURES



misc\_feature  
vector\_side:left  
28568..31135  
/notes="assembly\_fragment"  
misc\_feature  
31236..41487  
/notes="assembly\_fragment"  
misc\_feature  
41588..74365  
/notes="assembly\_fragment"  
misc\_feature  
74466..96541  
/notes="assembly\_fragment"  
misc\_feature  
96642..98613  
/notes="assembly\_fragment"  
clone end:SP6  
vector\_side:right

ORIGIN

Query Match 22.1%; Score 85; DB 2; Length 98613;  
Best Local Similarity 62.4%; Pred. No. 2.8e-08;  
Matches 133; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
QY 9 AGATCTTTGACATCTACGTGCACCGCTCACCTGACCTGACCTGCCCTAGGGCTGACGAGGATG 68  
Db 72330 AGATCTTTGAAATGATGTGACGGAGCCGACTGACCTGAGAGGGCGGAAATAAGGACA 72389  
QY 69 CCATCAGCTGCTGGGGAAGGCGAGTATGTGGAGGTCTCTGATGCGAGCCACCCACTCGCT 128  
Db 72390 GTGTGTCCTGACGAGGAGCGGCGAGTTGTGGAGGTCTTGATGCTTTTCATCTCTGAGAGAT 72449  
QY 129 GGCTTGTCCGACCAACCCACCAAGTCAGCCCTCAGCGAGGCTGGGTGTCACCAG 188  
Db 72450 GGCTGGTGAGAACCAAAACCCACCAACCAACCTGCTCAGACGGGTGGGTGCTGCCG 72509  
QY 189 CCTACCTGGACGAGGCTCAAGCTGCACCTG 221  
Db 72510 GGTACCTGGACGAGGAGGAGGAGGTCAGGCTG 72542

RESULT 14  
BX572602/c 349981 bp DNA linear BCT 18-DEC-2003  
LOCUS BX572602 BX571963  
DEFINITION Rhodopseudomonas palustris CGA009 complete genome; segment 10/16.  
ACCESSION BX572602  
VERSION BX572602.1 GI:39649689  
KEYWORDS complete genomes.  
SOURCE Rhodopseudomonas palustris CGA009  
ORGANISM Rhodopseudomonas palustris CGA009

REFERENCE 1 (bases 1 to 349981)  
AUTHORS Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M.L., Pelletier, D.A., Beatty, T.J., Lang, A.S., Tabita, F.R., Gibson, J.L., Hanson, T.E., Torres y Torres, J., Peres, C., Harrison, F.H., Gibson, J. and Harwood, C.S.  
TITLE Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 349981)  
Larimer, F.W. and Harwood, C.S.  
CONSTRM Rhodopseudomonas genome consortium  
TITLE Direct Submission  
JOURNAL Submitted (24-JUL-2003) Submitted on behalf of the Rhodopseudomonas genome consortium, the DOE Joint Genome Institute, Production Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA, and the Genome Analysis Group, Oak Ridge National Laboratory, 1060 Commerce Park Drive, Oak Ridge, TN 37831, USA; larimerf@ornl.gov

FEATURES  
1. 349981  
Location/Qualifiers  
/organism="Rhodopseudomonas palustris CGA009"  
/mol\_type="genomic DNA"  
/strain="CGA009"  
/db\_xref="taxon:258594"  
228..1463  
/locus\_tag="RPA2769"

CDS  
228..1463  
/locus\_tag="RPA2769"  
/function="InterPro IPR000160 COGs COG2199"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="putative diguanylate cyclase (GGDEF)"  
/protein\_id="CAE28211.1"  
/db\_xref="GI:39649690"  
/transl\_table=11  
/translation="MSLDISTFLVATLICALGVMHLHYFGRQKMPALNWNAYLL  
GATVALVTLSPFGEATITLMAISIGFIACGMMNAARVPHGLKPNWGVFLGALAW  
LAASLPPEQALMTIGAAIVAAVAGLTASQAAERKSMHRRWPTTVVFLHGA  
LMLPTVLADLPSPHSGFAGNVVVLFAVELVLAIGTVFVIMLVRSVSLHKTAA  
SLDPLTGMNRRGFESEACARMIEREAMAGRPVSALIPDIDHFKSINDRFHGPAGDEV  
KLPAIVTNSLRISDLSRIGGSEFAALLPCPIDENATADRVREAFQNCIGLEVDAP  
VATTVSGVAGGPAAVELEVLAAADTALYQKRGRRVATVVEQVPSLESRRKAA  
QTRHQVARGSEVQAVA"  
order(240..308,327..386,414..482,501..554,582..641,  
678..731,774..842)  
/locus\_tag="RPA2769"  
/note="7 probable transmembrane helices predicted  
by TMHMM2.0"  
1542..1844  
/locus\_tag="RPA2770"  
1542..1844  
/locus\_tag="RPA2770"  
/function="COGs COG2329"  
/note="observed by proteomics  
Citation: Proteomics from VerBerkmoes et al. (2003)  
unpublished"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="conserved unknown protein"  
/protein\_id="CAE28212.1"  
/db\_xref="GI:39649691"  
/translation="MITEIQIDVKPGSEADFEAAVAKARAFGRSKGFCGFLHKS  
IKPQRLNWKQTLNHTVDFRGSNFTWRGLVGFFAAPDVEHTETVLTSQA"  
complement(1828..2994)  
/locus\_tag="RPA2771"  
complement(1828..2994)  
/locus\_tag="RPA2771"  
/function="InterPro IPR002549 COGs COG0628"  
/note="observed by proteomics  
Citation: Proteomics from VerBerkmoes et al. (2003)  
unpublished"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="possible transport protein/permease"  
/protein\_id="CAE28213.1"  
/db\_xref="GI:39649692"  
/translation="MKGPVVTATLPDRAPSRDTLAWATLGGIATVSPAALLYAAWQF  
AGTLFLIAGVLFGVFLNALTELTGKLVGGPHGLRLALICLVLTGLTIGVAFLGNTI  
SQAKALSGTIKTOIVNVKALLERNGVDTGFLNPTPSAAHPETTTVPORNLPSAG  
ETASGSGAIVSQTKLILGTIGVGNFIVMFLGAPAAQSPVYVNGVAVVPRRYA  
QADVITDGENLKKWLAQLITMNAFAVATGSLIGIFGATLQAGLLAFIPT  
VGAIIIGLIIIVLAIGSGVIAVSFVLGVHLESYILTFIQRAIDITPET"  
order(2869..2934), complement(2797..2862),  
complement(2701..2766), complement(2401..2466),  
complement(2227..2292), complement(2122..2187),  
complement(2044..2109), complement(1939..2004)  
/locus\_tag="RPA2771"  
/note="8 probable transmembrane helices predicted  
by TMHMM2.0"  
3241..4188  
/gene="rlpA2"  
/locus\_tag="RPA2772"  
3241..4188  
/gene="rlpA2"  
/locus\_tag="RPA2772"  
/function="InterPro IPR005132 COGs COG0797"



Shiba,T., Sakaki,Y., Hattori,M. and Omura,S.  
Complete genome sequence and comparative analysis of the industrial  
microorganism Streptomyces avermitilis  
Nat. Biotechnol. 21 (5), 526-531 (2003)  
22608306  
PUBMED 12693562  
3 (bases 1 to 300425)  
Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,  
Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osuno,T.,  
Kushida,N., Director-General of Biotechnology Center, Shiba,T.,  
Sakaki,Y. and Hattori,M.  
Direct Submission  
Submitted (29-MAR-2002) Director-General of Biotechnology Center,  
National Institute of Technology and Evaluation, Biotechnology  
Center, 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
(E-mail:bioente.go.jp, URL:http://www.bio.nite.go.jp/,  
Tel:81-3-3481-1933, Fax:81-3-3481-8424)  
This work was done in collaboration with Haruo Ikeda(\*1), Jun  
Ishikawa(\*2), Akiharu Hanamoto(\*3), Chigusa Takahashi(\*3), Mayumi  
Shinose(\*3), Hiroshi Horikawa(\*4), Hidekazu Nakazawa(\*4), Tomomi  
Osuno(\*4), Norihiro Kushida(\*4), Hisashi Kikuchi(\*4), Tadayoshi  
Shiba(\*5), Yoshiyuki Sakaki(\*6,\*7), Masahira Hattori(\*1,\*7)  
and Satoshi Omura(\*1,\*3).  
Final finishing process and all annotation were done by H. Ikeda  
and J. Ishikawa.  
\*1 Kitasato Institute for Life Sciences, Kitasato University  
\*2 National Institute of Infectious Diseases  
\*3 The Kitasato Institute  
\*4 National Institute of Technology and Evaluation  
\*5 School of Science, Kitasato University  
\*6 Institute of Medical Science, University of Tokyo  
\*7 RIKEN, Genomic Sciences Center  
Following url is also available.  
http://avermitilis.lsl.kitasato-u.ac.jp.  
Location/Qualifiers  
1..300425  
/organism="Streptomyces avermitilis MA-4680"  
/mol\_type="genomic DNA"  
/strain="MA-4680"  
/db\_xref="taxon:227882"  
/note="This strain is also named as strain: ATCC 31267,  
NCIMB 12894 or NRRL 8165.-synonym: Streptomyces  
avermitilinus"  
complement (382..645)  
/note="SAV258"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC67967.1"  
/db\_xref="GI:29603895"  
/translation="MMFALAPEVSVTDTEHGVLLDERTGRYLLNGTGALALRLL  
DGTSPQGVASLCORFPDAGTAQVEQVDNLLALRTAKVAAS"  
complement (642..2552)  
/note="SAV259"  
/codon\_start=1  
/transl\_table=1  
/product="putative membrane-associated Zn-dependent  
protease"  
/protein\_id="BAC67968.1"  
/db\_xref="GI:29603896"  
/translation="MPAVRLPHLLTLVEGTGMEFLVLPDPAGAMALAMPAGPPP  
THVLEHSGRWIVGCNNEEDVMVTAGPRHVALGRTADADVLRRLDVSVDRL  
DGLAATAGSPHLVASNGEVRVQGTLSAACQIYATVAGTVLADSPQALADRIHAD  
IDEPVALALSPFPGLPWSLWGVCSPTTGCHLEIGRDGTQRQWMTAPDP  
RPLGNMIRQALAEAVATCTAGQGVVSADLSGLDSTSLCFADADRHADHLTVHY  
QALGRSDDRYWARQAEALPCARHLLVGQCAPDVPYAPLAHVDREGPLTFVRSRA  
TTEHLARLAAGSRHRLQIGIGGDELPHSPMALHVLVRHPLAALPHLARLSMRW  
PLTTARTLASPSYTLWLACSFPADRDHWDGADWEPKLPKPPATPHAAATVR  
QLALAAATPQPLATPLVQHEMIRLTRINGNATRRNSHIGHTGVTHAFPTDVRLE  
AILAVLRDEDRVALGQFKPLGLTLRSIVPAGLLSRTTKGASAEIYAGLRRLGELLK  
LMDSHLVMEGFVDADVLRVLRGLHADYTRLEIMFLDSTLACELWRLTSPADTARATA  
VVPSPRTVKEI"  
3174..3572

/note="SAV260"  
/codon\_start=1  
/transl\_table=1  
/product="putative transposase"  
/protein\_id="BAC67969.1"  
/db\_xref="GI:29603897"  
/translation="MGRAGPVAFILTDORRNYGTNOVDDQGLAGYFLFLDRDARRRA  
MTCGRASQLGYAVQLCTVRELCTFLDNPEPAEYVAYVAEQLGHPGSLVLAGYGAER  
TRWDHQIAIKDAIGTRDLKGDANWKLADGD"  
3604..3834  
/note="SAV261"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC67970.1"  
/db\_xref="GI:29603898"  
/translation="MLGGQGVPLDLMIEGESSQMSERDALRGLRALMWVVAIEVPV  
YLVMLVEVSYLTGQQLPRGTBEEDKPRADTD"  
4027..4500  
/note="SAV262"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC67971.1"  
/db\_xref="GI:29603899"  
/translation="MQADAMAVAVMELQGAANRLWEGFAERGRTEFLTLAFAG  
AARARIAGTQVQSLGVGQGAELLSSERVASKOTVAAREPLAQAAATAAPLMRCS  
APAVVATEQLNALADVENTARLEALEAFGRVNVATPRLSWARRAVSSG"  
complement (4811..6445)  
/note="IS6 family; completely same as SAV7545 and SAV75480  
SAV263"  
/codon\_start=1  
/transl\_table=1  
/product="putative IS6 family transposase"  
/protein\_id="BAC67972.1"  
/db\_xref="GI:29603900"  
/translation="MILAECELLTMEFHEGLVVEVSPGGVHLHVARTVESVPV  
PCPDATFVRHSGYQRRLDAGVGGQVSIETLVRLFCDDQCARVFAEQVDGL  
TVYGRTRPQRLCSAIJALAGRAGERLAARLPVPSVTRTLTALVNALPQYAEETP  
RVLGVDEPATRKHKYGTVLVDCETHAPIDILIDREESATFAAWLTHFGEVEICDRG  
GAFADGARTGAPDAVOVADLHLWHNLAEITVKSLSVSHSSCLREPALASEPPEVLHP  
PLSHAGRLAERARRHAAVHDLLGGLGTLIRAVARLELRNTVRYAVADYWEELATG  
RWQNLPSTLDPYKSYLHQWHEGHTSGAKLHAELEERGETGTSYSVVRYLRRFRFSTG  
DRPPARPFGVRKVTGWTIRPNDRMDDDOQKLKAILARCPELEAATGHVRSFAAMMA  
IRASRLPEWATARADBEHGLRGFADGLLADLDAVILGLSTWSSGCGVEGVTDIKL  
LKRQAGRAGLPLLRKRVLLVAADRQRHRTVNTQTAH"  
complement (6592..8802)  
/note="SAV264"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC67973.1"  
/db\_xref="GI:29603901"  
/translation="MTPATPTPPSTPPSPQCGHGTTPEDPAGRGIRVSGHTACLA  
HLSADADRDAYLAGLSFGNNVDHRTGSFQGLDLDLADLRDPTDGPRLGVARFRSAT  
FORDYFRSATQKWADPQSATFOGNAGFRLATFEGDAGFGSARFGDAGFGSARFEG  
DASFVATFEGDASFVATFEGDASFVATFQSAASLGNWAFQGDVKFESATFQSDAV  
FGADFRGDARFGLAARFNARFGLARFNGIAFSLVTFHGDANFRSAIFRGDASFVS  
ATFQSDTGLEAATFSDHTRFDRAFTFGDARFVEATFRGTSFVSATLQSNANFLGVTF  
ERDAKFESIQQCDATFVSATFECDAFCFQVGFERSASGLPLVCAGRVKLSGAVFGCP  
VTLFARCLCECRTRWSTABLRVATVDFAHAVFEYPLTIAAADPFLVJHNGRPV  
AQGLSDIADPIEVRIASLRGVDAHLVSDVDSGLCLFTGVHLQIIELEGTCFSASV  
PRTTHWRPERLFTCRBALAEEGFWRASQPTPVFGMNAAVAGAVHAGPAQLAPVRA  
LKAATEDSRNECAADFFYGENEMERHRTGSTRTERGLLHGVLLSGVGLASRAIG  
WFEAAWLVTLVLMGFGLPKQSPKQVATGTPPFGCKKTLIEDKADPQNPTGDRTISE  
RFEKALTGLTNSVWFRSSQDLTTAGTVIEMASRVTEPVLVLAVLAVNRVRK"  
9294..9707  
/note="SAV265"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC67974.1"  
/db\_xref="GI:29603902"

/translation="MGTRLRLRRADFSRGRKLTOKVPVSRPQGVQGLTNTYLSP  
 TEYDILLASLPAAVLSTKFSVPPLGVDFDPLQGLVLAEEFTTDEETQSFIPPSEC  
 VAETDDARFTGGRLVETSRQELLGLAEYGLRPQ"  
 complement(9955..10614)  
 /note="SAV266"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC67975.1"  
 /db\_xref="GI:29603903"  
 /translation="MPHVLRKRESSCSYLNRLLADANRLPVQFRPILGHNHHRD  
 DNRGRWTPPTCLSLTGTAEALSHAI PALAVIGEPFAENRLQSLSGQAVEVLRFP  
 ACRCMARFGHGLVVRTIPPEAVCLRHQRWLLGDEQYRLHALPELQANRRHCLV  
 RRTRASILAREYAVARDCLRDWFQDGKPSALRQWNKRIQLGDEPVGDETPVAPPS  
 G"

CDS

10830..11180  
 /note="SAV267"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC67976.1"  
 /db\_xref="GI:29603904"  
 /translation="MLEDPAAHGVDLCTNVLHETLGDENPATRAHEEFVLPHLRH  
 RVLVQVARASRSLEITVIDDSRQPRIVERGPWALWDEYESGTVPQGGIRLCLSLH  
 AKGNRWPLSPTTC"  
 complement(11530..11979)  
 /note="SAV268"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC67977.1"  
 /db\_xref="GI:29603905"

CDS

Query Match 12.3%; Score 47.4; DB 1; Length 300425;  
 Best Local Similarity 49.0%; Pred. No. 2;  
 Matches 126; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy	18	ACATCTACGTGGTCACCGCTGACTACCTGCCCTAGGGGTGAGCAGGATGCCATCACGC	77
Db	171375	ACATCGACCTCGAGCGCGGACCGACTGCCGTCGGGCAGTTACGTCGTCTACCTCAAGG	171434
Qy	78	TGCGGGAAGCGCAGTATGTGGAGGTCTCTGGATGCAGCCACCCACTGCCGTGGCTTGTC	137
Db	171435	TCTGGGAACGCCAGGTCAACGAGGTGAGGACCCCTCGCTCCACGAAACCGCGCTCGGC	171494
Qy	138	GCACCAAGCCACCAAGTCCAGCCCTCAGCGGAGGGCTGGGTGTCCACAGCCTACCTGG	197
Db	171495	TGCACCAAGCCCGACTCGCGGAGCGCGCCAGGTGGTGTGGCAGGTCCCGGCACCTGCCG	171554
Qy	198	ACAGGAGGCTCAAGCTGTACCTGAGTGGGGGGCGCTGAGGCCCTGAGTTCCCTGGGG	257
Db	171555	TCAAGAGCCGGCAGACCCCGTGCACCGGATCCCGCCGAGCAGCACCGGACCGGGGCA	171614
Qy	258	AGCTGTGTCTGAAGAC	274
Db	171615	CGGCGGTCTCTGACGTC	171631

Search completed: September 19, 2004, 19:27:13  
 Job time : 1576.62 secs











```
/gene="OBSCN"
/number=41
11931..12465
/gene="OBSCN"
/number=41
12466..12729
/gene="OBSCN"
/number=42
12730..13093
/gene="OBSCN"
/number=42
13094..13357
/gene="OBSCN"
/number=43
13358..16767
/gene="OBSCN"
/number=43
16768..17031
/gene="OBSCN"
/number=44
17032..17572
/gene="OBSCN"
/number=44
17573..17836
/gene="OBSCN"
/number=45
17837..18195
/gene="OBSCN"
/number=45
18196..18459
/gene="OBSCN"
/number=45
18460..21988
/gene="OBSCN"
/number=45
21989..22252
/gene="OBSCN"
/number=47
22253..22616
/gene="OBSCN"
/number=47
22617..22880
/gene="OBSCN"
/number=48
22881..23475
/gene="OBSCN"
/number=48
23476..23700
/gene="OBSCN"
/number=49
23701..24685
/gene="OBSCN"

Query Match      67.8%; Score 261; DB 9; Length 34667;
Best Local Similarity 100.0%; Pred. No. 7.3e-50;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 125  GCCCTGCCAGCCCAAGTTCACAGAGGCTCTGAGGAATGAAGAGGCCGTGGAGGGGCCACA 184
Db 10211 GCCCTGCCAGCCCAAGTTCACAGAGGCTCTGAGGAATGAAGAGGCCGTGGAGGGGCCACA 10270

Qy 185  GCCATGTTGTTGTTGTAACCTGAGCAAGTGGCCCTGTGGAGTGGAGGAGGGCCCGAG 244
Db 10271 GCCATGTTGTTGTTGTAACCTGAGCAAGTGGCCCTGTGGAGTGGAGGAGGGCCCGAG 10330

Qy 245  AACCTCAGAGATGGGGACAGATACATCTTGAGGCAGAGGGGACCAAGTGTGAGCTGCAG 304
Db 10331 AACCTCAGAGATGGGGACAGATACATCTTGAGGCAGAGGGGACCAAGTGTGAGCTGCAG 10390

Qy 305  ATCTGTGCTGCTGCGGACGCGGGAGTACTTGTGTGTCGGGCGAGGAGG 364
Db 10391 ATCTGTGCTGCTGCGGACGCGGGAGTACTTGTGTGTCGGGCGAGGAGG 10450

Qy 365  ACCTCAGCCAGCTCACCATC 385
Db 10451 ACCTCAGCCAGCTCACCATC 10471

RESULT 4
AL353593      135964 bp      DNA      linear      PRI 13-FEB-2002
LOCUS      Human DNA sequence from clone RP5-1139B12 on chromosome 1q42.1-43,
DEFINITION      complete sequence.
ACCESSION      AL353593
VERSION      AL353593.33 GI:18673899
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 135964)
AUTHORS      Almeida,J.
TITLE      Direct Submission
JOURNAL      Submitted (13-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonesrequest@sanger.ac.uk
On Feb 14, 2002 this sequence version replaced gi:17977879.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Emi, EMBL; Swi,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP5-1139B12 is from the library RPCI-5 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pcYPAC2
This sequence is the entire insert of clone RP5-1139B12.
FEATURES
Location/Qualifiers
1..135964
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="q42.1-43"
/clone="RP5-1139B12"
/clone_lib="RPCI-5"
25928..26054
/misc_feature
/note="Sequence from overlapping clone RP11-520H14
(AL359510). Assembly confirmed by restriction digest."

ORIGIN

Query Match      67.8%; Score 261; DB 9; Length 135964;
Best Local Similarity 100.0%; Pred. No. 5.7e-50;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 125  GCCCTGCCAGCCCAAGTTCACAGAGGCTCTGAGGAATGAAGAGGCCGTGGAGGGGCCACA 184
Db 79224 GCCCTGCCAGCCCAAGTTCACAGAGGCTCTGAGGAATGAAGAGGCCGTGGAGGGGCCACA 79283
```

```

Qy 185 GCCATGTTGGTGTGAACAGAGTGGCCCTGTGGAGTGGAGGAAGGGCCCGAG 244
Db 79284 GCCATGTTGGTGTGAACAGAGTGGCCCTGTGGAGTGGAGGAAGGGCCCGAG 79343
Qy 245 AACCTCAGAGATGGGACAGATACATCCTCAGGAGGAGGGACAGGTTGTGAGCTGCAG 304
Db 79344 AACCTCAGAGATGGGACAGATACATCCTCAGGAGGAGGGACAGGTTGTGAGCTGCAG 79403
Qy 305 ATCTGTGCTGCTGCCATGCGGAGCGCGGAGTACTTGTGTGTGTCGGCGCAGGAGAGG 364
Db 79404 ATCTGTGCTGCTGCCATGCGGAGCGCGGAGTACTTGTGTGTGTCGGCGCAGGAGAGG 79463
Qy 365 ACCTCAGCCAGCTCACCATC 385
Db 79464 ACCTCAGCCAGCTCACCATC 79484

RESULT 5
LOCUS AC026657 164766 bp DNA linear HTG 01-SEP-2000
DEFINITION Homo sapiens chromosome 1 clone RP11-245P10, WORKING DRAFT
ACCESSION AC026657
VERSION AC026657.4 GI:9958202
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
The sequence of Homo sapiens clone
2 (bases 1 to 164766)
Waterston,R.H.
Direct Submission
Submitted (22-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:7637349.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0245P10
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 139884 bases at least Q40
Consensus quality: 147686 bases at least Q30
Consensus quality: 151469 bases at least Q20
Insert size: 169000; agarose-fp
Insert size: 161074; sum-of-contigs
Quality coverage: 3.60 in Q20 bases; agarose-fp
Quality coverage: 3.92 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1124: contig of 1124 bp in length
* 1125 1224: gap of unknown length
* 1225 3032: contig of 1808 bp in length
* 3033 3132: gap of unknown length

FEATURES
source
1..164766
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-245P10"
1..1124
/note="assembly_name:Contig14"
1225..3032
/note="assembly_name:Contig21"
3133..4493
/note="assembly_name:Contig23"
4594..5860
/note="assembly_name:Contig24"
5961..7670

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

```



```

* 34916 45742: contig of 10827 bp in length
* 45743 45842: gap of unknown length
* 45843 62987: contig of 17145 bp in length
* 62988 63087: gap of unknown length
* 63088 87584: contig of 24497 bp in length
* 87585 87684: gap of unknown length
* 87685 114765: contig of 27081 bp in length
* 114766 114865: gap of unknown length
* 114866 142334: contig of 27469 bp in length
* 142335 142435: gap of unknown length
* 142435 174612: contig of 32178 bp in length.

FEATURES
    source
        Location/Qualifiers
            1..174612
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="1"
                /clone="RP11-661B12"
            1..2478
                /note="assembly_name:Contig4"
            2579..6948
                /note="assembly_name:Contig5"
            7049..9373
                /note="assembly_name:Contig6"
            9474..13303
                /note="assembly_name:Contig7"
            13404..18270
                /note="assembly_name:Contig8"
            18371..25589
                /note="assembly_name:Contig9"
            25690..34815
                /note="assembly_name:Contig10"
            34916..45742
                /note="assembly_name:Contig11"
            45843..62987
                /note="assembly_name:Contig12"
            clone_end:SP6
            vector_side:right
            63088..87584
                /note="assembly_name:Contig13"
            87685..114785
                /note="assembly_name:Contig14"
            114866..142334
                /note="assembly_name:Contig15"
            142435..174612
                /note="assembly_name:Contig16"

ORIGIN
Query Match      67.8%; Score 261; DB 2; Length 174612;
Best Local Similarity 100.0%; Pred. No. 5.5e-50;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 GCCTGCGCAGCAAGTTTCACAGAGGCTCAGGATCAGAGCCGCTGGAGGGGCCACA 184
    |||
Db 170271 GCCTGCGCAGCAAGTTTCACAGAGGCTCAGGATCAGAGCCGCTGGAGGGGCCACA 170212

QY 185 GCCATGTTGTGTGTGAACCTGAGCAAGTGGCCCTGTGTGAGTGGAGGAAGGGGCCCGAG 244
    |||
Db 170211 GCCATGTTGTGTGTGAACCTGAGCAAGTGGCCCTGTGTGAGTGGAGGAAGGGGCCCGAG 170152

QY 245 AACCTCAGATGGGACATACATCCTCAGCAGAGGGGACCCAGGTGTGAGCTGCAG 304
    |||
Db 170151 AACCTCAGATGGGACATACATCCTCAGCAGAGGGGACCCAGGTGTGAGCTGCAG 170092

QY 305 ATCTGTGGCTGGCCATGCGGAGCGCCGGGGAGTACTTGTGTGTGTGCGGGCAGGAGG 364
    |||
Db 170091 ATCTGTGGCTGGCCATGCGGAGCGCCGGGGAGTACTTGTGTGTGTGCGGGCAGGAGG 170032

QY 365 ACCTCAGCAGGCTCACCATC 385
    |||
Db 170031 ACCTCAGCAGGCTCACCATC 170011

```

```

RESULT 7
LOCUS AX069754/c 352 bp DNA linear PAT 25-JAN-2001
DEFINITION Sequence 226 from Patent WO0102568.
ACCESSION AX069754
VERSION AX069754.1 GI:12579539
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
    1 Williams, L.T., Escobedo, J., Innis, M.A., Garcia, P.D., Klinger, J.,
      Kassam, A., Reinhard, C., Randazzo, F., Kennedy, G.C., Pot, D.,
      Lamson, G., Drmanac, R., Crkenjakov, R., Drmanac, S., Dickson, M.,
      Labat, I., Leshkowitz, D., Kita, D., Garcia, V. and Strache-Crain, B.
      Human genes and gene expression products
      Patent: WO 0102568-A 226 11-JAN-2001;
      CHIRON CORPORATION (US); HYSEQ, INC. (US)
FEATURES
    source
        Location/Qualifiers
            1..352
                /organism="Homo sapiens"
                /mol_type="unassigned DNA"
                /db_xref="taxon:9606"
            1..352
                /note="n = A, T, C or G"
            misc_feature
                1..352

ORIGIN
Query Match      65.9%; Score 253.6; DB 6; Length 352;
Best Local Similarity 86.4%; Pred. No. 8.1e-48;
Matches 280; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 62 GATGCTGGGAGTACTCTCTGTGTGTGTGGGAGGAGGAGGACCTTGCCACTCTCACCGTG 121
    |||
Db 342 GACGCTGGGAGTACTCTGTGTGTGTGTGGGAGGAGGAGGACCTTGCCACTCTCACCGTC 283

QY 122 AAGGCCCTGCCAGCAAGTTCACAGAGGCTCAGGAAATGAAGAGGCGCTGGAGGGGCC 181
    |||
Db 282 AGGCCCATGCTTCNAGTTCATAGAGGCTCAGGAAATGAAGAGGCGCACAGAGGGGAC 223

QY 182 ACAGCCATGTTGTGTGTGAACCTGAGCAAGTGGCCCTGTGTGAGTGGAGGAAGGGGCC 241
    |||
Db 222 ACGGCCACGCTGTGTGTGTGAGCTGAGCAAGGCGGCACCGGTGGAGTGGAGGAGGGCAT 163

QY 242 GAGAACCTCAGAGATGGGRCAGATACATCTCTGAGGAGGAGGAGGAGGAGTGTGAGCTG 301
    |||
Db 162 GAGACCTCAGAGATGGGRCAGACACAGCTTGAAGGAGGAGTGGTCCAGTGTGAGCTG 103

QY 302 CAGATCTGTGCCCTGGCCATGGCGGACGCCGGGAGTACTTGTGTGTGCGGGCAGGAG 361
    |||
Db 102 CAGATCGTGGCTGGCTGTGTGTGTGGATGCCGGGAGTACTCTGTGTGTGCGGGCAGGAG 43

QY 362 AGGACCTCAGCAGCGCTCACCATC 385
    |||
Db 42 AGGACCTCAGCAGCGCTCACCATC 19

RESULT 8
LOCUS AL645854/c 211829 bp DNA linear ROD 05-APR-2002
DEFINITION Mouse DNA sequence from clone RP23-192P17 on chromosome 11,
complete sequence.
ACCESSION AL645854
VERSION AL645854.10 GI:20068636
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
    1 Smith, M.
      Direct Submission
      Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,

```

Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
On Apr 7, 2002 this sequence version replaced gi:19699553.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e. phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep/

Best Local Similarity 50.9%; Score 196; DB 10; Length 211829;  
Matches 229; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
For further details see http://www.chori.org/bacpac/home.htm  
VECTOR: PBAC3.6  
Location/Qualifiers  
1. .211829  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="11"  
/clone="RP23-192P17"  
/clone\_lib="RPCI-23"

FEATURES  
source  
1. .211829  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="11"  
/clone="RP23-192P17"  
/clone\_lib="RPCI-23"

ORIGIN  
Query Match 50.9%; Score 196; DB 10; Length 211829;  
Best Local Similarity 80.6%; Pred. No. 4e-35;  
Matches 229; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
102 CCTGTGCCACTCTACCGTGAAGCGCTCCAGCCAGATTCACAGAGGCTCTGAGGAATG 161  
Db 46277 CCTGTGCCACTCTACCGTGAAGCGCTCCAGCCAGATTCACAGAGGCTCTGAGGAATG 46218  
162 AAGAGGCGCTGAAGGGGCCACAGCCATGTTGTGTGTAAGTCCAGCAAGTGTGCCCTCG 221  
Db 46217 AAGAGGCGCTGAAGGGGCCACAGCCATGTTGTGTGTAAGTCCAGCAAGTGTGCCCTCG 46158  
222 TGGAGTGAAGAGGGGCCAGAACCTTCAGAGATGGGGACAGATACATCTCTGAGGCGAG 281  
Db 46157 TGGAGTGAAGAGGGGCCAGAACCTTCAGAGATGGGGACAGATACATCTCTGAGGCGAG 46098  
282 AGGGACCAAGTGTGAGCTGCAGATCTGTGCTGGCCATGGCGGAGCGCGGGAGTACT 341  
Db 46097 ATGGGGCCATGTTGTGAGCTGCAGATCTGTGCTGGCCATGGCGGAGTACT 46038  
342 TGTGTGTGTGGGGCAGGAGGAGGACCTTCAGCCAGCTCAACATC 385  
Db 46037 CATGTGTGTGGGGCAGGAGGAGGACCTTCAGCCAGCTCAACATC 45994

RESULT 9  
AC026657/c  
LOCUS  
DEFINITION  
Homo sapiens chromosome 1 clone RP11-245P10, WORKING DRAFT  
SEQUENCE, 31 unordered pieces.  
AC026657.4 GI:9958202  
KEYWORDS  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 164766)

Waterston,R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 164766)  
Waterston,R.H.  
Direct Submission  
Submitted (22-MAR-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Sep 1, 2000 this sequence version replaced gi:7637349.

Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H NH0245P10  
----- Summary Statistics -----  
Sequencing vector: plasmid, 0%  
Chemistry: Dye-primer Et; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 139884 bases at least Q40  
Consensus quality: 147686 bases at least Q30  
Consensus quality: 151469 bases at least Q20  
Insert size: 169000; agarose-fp  
Insert size: 161074; sum-of-contigs  
Quality coverage: 3.60 in Q20 bases; agarose-fp  
Quality coverage: 3.92 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 31 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1124: contig of 1124 bp in length  
\* 1125 1224: gap of unknown length  
\* 1225 3032: contig of 1808 bp in length  
\* 3033 3132: gap of unknown length  
\* 3133 4493: contig of 1361 bp in length  
\* 4494 4593: gap of unknown length  
\* 4594 5860: contig of 1267 bp in length  
\* 5861 5960: gap of unknown length  
\* 5961 7670: contig of 1710 bp in length  
\* 7671 9669: contig of 1899 bp in length  
\* 9670 11715: contig of 1946 bp in length  
\* 11716 11815: gap of unknown length  
\* 11816 14243: contig of 2428 bp in length  
\* 14244 14343: gap of unknown length  
\* 14344 16887: contig of 2544 bp in length  
\* 16888 16987: gap of unknown length  
\* 16988 19248: contig of 2260 bp in length  
\* 19249 19348: gap of unknown length  
\* 19349 21376: contig of 2028 bp in length  
\* 21377 21475: gap of unknown length  
\* 21476 25025: contig of 3550 bp in length  
\* 25026 25125: gap of unknown length  
\* 25126 28051: contig of 2926 bp in length  
\* 28052 32054: contig of 3903 bp in length  
\* 32055 32154: gap of unknown length  
\* 32155 35716: contig of 3562 bp in length  
\* 35717 35816: gap of unknown length  
\* 35817 40891: contig of 5075 bp in length  
\* 40892 44027: gap of unknown length  
\* 44028 44127: gap of unknown length

* 44128	48899:	config	of 4772 bp in length
* 48900	48999:	gap	of unknown length
* 49000	54255:	config	of 5256 bp in length
* 54256	54355:	gap	of unknown length
* 54355	55994:	config	of 4639 bp in length
* 55995	55094:	gap	of unknown length
* 55095	64390:	config	of 5296 bp in length
* 64391	64490:	gap	of unknown length
* 64491	70865:	config	of 6375 bp in length
* 70866	70965:	gap	of unknown length
* 70966	78667:	config	of 7702 bp in length
* 78668	78767:	gap	of unknown length
* 78768	87207:	config	of 8440 bp in length
* 87208	87307:	gap	of unknown length
* 87308	97858:	config	of 10551 bp in length
* 97859	97958:	gap	of unknown length
* 97959	109520:	config	of 11562 bp in length
* 109521	109620:	gap	of unknown length
* 109621	120908:	config	of 11288 bp in length
* 120909	121008:	gap	of unknown length
* 121009	141477:	config	of 20469 bp in length
* 141478	141577:	gap	of unknown length
* 141578	162672:	config	of 21095 bp in length
* 162673	162777:	gap	of unknown length
* 162778	163974:	config	of 1202 bp in length
* 163975	164074:	gap	of unknown length
* 164075	164766:	config	of 692 bp in length.

## ORIGIN

Query Match	49.9%;	Score 192.2;	DB 2;	Length 164766;
Best Local Similarity	83.5%;	Pred. No. 3.1e-34;		
Matches 218;	Conservative 0;	Mismatches 43;	Indels 0;	Gaps 0;
QY 125	GCCTGCCAGCAAGTTCACAGAGGGTCTGAGGAATCAAGAGGCGCTGGAAAGGGGCCACA	184		
DB 133606	GCCTGCCGGCCAGATTCATAGAAGATGTGAGAAATCACAGAGCCACAGAGAGGGGCCACA	133547		
QY 185	GCCATGTTGTGGTGTGAACCTAGCAAGGTGGCCCTGTGGAGTGGAGGAGGGGCCCGAG	244		
DB 133546	GCTGTGCTGCACTGTGAGCTGAGCAAGCGGCCGCCCTGTGGTGGCGGAGGGGGTCTGAG	133487		
QY 245	AACCTCAGAGATGGGGACAGATACATCCTCAGGCGAGAGGGGACCAAGTGTGTGAGCTGCAG	304		
DB 133486	ACCCTCAGAGATGGGGACAGATATAGCCTGAGGCGAGACGGGACGAGGTGTGTGAGCTGCAG	133427		
QY 305	ATCTGTGCGCTGGCCATGGCGGACGCGGGGAGTACTTGTGTGTGTGTGCGGGCAGAGGAGG	364		
DB 133426	ATTCGTGCGCTGGCTGTGGAGGACACTGGAGAGTATTTGTGTGTGCGGGCAGGAGAGA	133367		
QY 365	ACCTCAGCCAGCGCTCACCATC	385		
DB 133366	ACCTCAGCTACACTCACTGC	133346		

RESULT 10  
AX463538

LOCUS	AX463538	2155 bp	DNA	linear	PAT 15-JUL-2002
DEFINITION	Sequence	102 from Patent WO0248337.			
ACCESSION	AX463538				
VERSION	AX463538.1	GI:21866312			
KEYWORDS					

REFERENCE  
AUTHORS

## ORIGIN

Query Match 49.5%; Score 190.6; DB 6; Length 2155;  
 Best Local Similarity 83.1%; Pred. No. 1.5e-33;  
 Matches 217; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 125 GCCCTGGCAGCAAGTTCACAGAGGGTCTGAGGAATGAAGAGGCGGTGGGAAGGGGCCACA 184  
 Db 205 GCCCTGGCAGCAAGTTCACAGAGGGTCTGAGGAATGAAGAGGCGGTGGGAAGGGGCCACA 264

Qy 185 GCCATGTGTGGTGTGAAGTTCAGCAAGTGGCCCTCTGGAGTGGAGGAGGGCCCGAG 244  
 Db 265 GCCACACTGCAATGTGAGCTGAGCAAGGACGCCCTCTGGAGTGGAGGAGGAGGCTTGG 324

Qy 245 AACCTCAGAGATGGGGACAGATACATCTCAGGAGGAGGAGGACCGAGTGTGAGCTGCAG 304  
 Db 325 GCTCTCAGAGATGGGGACAAATACAGCTGAGACACGCGGCTGTGTGAGCTGCAG 384

Qy 305 ATCTGTGGCTGGCAGTGGCGGAGCGCGGGAGTACTGTGTGTGCGGGCAGGAGG 364  
 Db 385 ATTCATGGCTGGTATGGCAGATACGCGGTGTACTCATGTGTGTGCGGAGGAGG 444

Qy 365 ACCTCAGCAGCTCACCATC 385  
 Db 445 ACCTCAGCTACCTCACTGTC 465

## RESULT 11

AC099089 260998 bp DNA linear HTG 10-MAY-2003  
 LOCUS Rattus norvegicus clone CH230-154E3, WORKING DRAFT SEQUENCE.  
 AC099089  
 DEFINITION HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
 VERSION AC099089.5 GI:30522701  
 KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE Rattus norvegicus  
 ORGANISM Rattus norvegicus

## REFERENCE

1 (bases 1 to 260998)  
 Murny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,  
 Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,  
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Chen,A.,  
 Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,X., Chen,Z., Chu,J.,  
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
 Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,  
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorensuhea,L., Loulsec,H., Lozada,R.J., Lu,X., Ma,J.,  
 Matheswar,M., Mahindratne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mathiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munida,M., Murphy,M., Nair,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwaokemele,O., Okwono,G., Olarpunsagoon,A., Pal,S., Parks,K.,  
 Pasceriak,S., Paul,H., Perez,A., Perez,E., Pfannkuch,C.,  
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,

Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sibson,I., Sitter,C.D., Smajs,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,B., Umani,K.,  
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,  
 Wright,D., Wrighe,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstein,G. and Gibbs,R.A.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Unpublished  
 2 (bases 1 to 260998)  
 Worley,K.C.  
 Direct Submission  
 Submitted (09-NOV-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 260998)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On May 10, 2003 this sequence version replaced gi:22855456.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GGLZ  
 Center clone name: CH230-154E3  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0;  
 Consensus quality: 216929 bases at least Q40  
 Consensus quality: 219228 bases at least Q30  
 Consensus quality: 220845 bases at least Q20  
 Estimated insert size: 226118; sum-of-contigs estimation  
 Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 260998: contig of 260998 bp in length.

## FEATURES

## source

1..260998 Location/Qualifiers  
 /organism="Rattus norvegicus"

```

/misc_feature
1..1486
/note="wgs_contig"

ORIGIN
Query Match 47.8%; Score 184.2; DB 2; Length 260998;
Best Local Similarity 81.6%; Pred. No. 1.9e-32;
Matches 213; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 125 GCCCTGCCAGCCAGTTTCACAGAGGCTCTGAGGAATGAGAGGCGCGTGAAGGGCCACCA 184
Db 161277 GCCCTGCCAGCCAGTTTCACAGAGGCTCTGAGGAATGAGAGGCGCGTGAAGGGCCACCA 161218

QY 185 GCCATGTTGTGGTGAAGTCAAGAGTGGCCCTCTGAGAGTGGAGAAAGGGCCCGAG 244
Db 161217 GTGACTCTGAGCTGCCAGATGAGCAAGGTGACCCCTCTGAGAGTGGAGAAAGGGGTGAGAG 161158

QY 245 AACCTCAGATGGGACAGATACATCTGAGGAGGAGGAGGACCGAGTGTGAGCTGCAG 304
Db 161157 ACCCTGAGAGATGGGGGAGATACAGCCCTGAGGAGGAGTGGGGCCGTGTGAGCTGCAG 161098

QY 305 ATCTGTGGCCCTGGCCATGGCGGAGCGCGGGAGTACTTGTGTGTCGGGCGGAGAGG 364
Db 161097 ATCCGTGGCCCTGATTCTGGAAGATGCTGGGAGTACTCATGCTGTGTGGGCGAGAGAG 161038

QY 365 ACCTCAGCAGCTCACCATC 385
Db 161037 ACGTCAGCCACACTGAGTGTC 161017

RESULT 12
HSM800639 HSM800639 2295 bp mRNA linear PRI 18-FEB-2000
LOCUS Homo sapiens mRNA; cDNA DKFP2p434C153 (from clone DKFP2p434C153);
DEFINITION partial cds.
ACCESSION AL080128
VERSION AL080128.1 GI:5262565
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Wambutt, R., Heubner, D., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
JOURNAL Direct Submission
COMMENT Submitted (16-JUN-1999) MIPS, Am Klopferstr. 18a, D-82152
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFP2p434C153) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
Location/Qualifiers
1..2295
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFP2p434C153"
/tissue_type="testis"
/clone_lib="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev stage="adult"
1..2295
/gene="DKFP2p434C153"
<1..348
/gene="DKFP2p434C153"
/note="similarity to titin"

gene
CDS

```

```

/codon_start=1
/product="hypothetical protein"
/protein_id="CAB45726.1"
/db_xref="GI:5262566"
/db_xref="SPTREMBL:O5Y485"
/translation="AMFCPKPPDPSLMISVTWVSFSLPARPIEDMNRQKATGATVTT
LQCKLRKAAPVWRKNTLKDGRYSLSKDGTSCEIQIRGLVITADAGEYSCICEQER
TSATLTIVRGKDHM"
polyA_signal
2232..2237
/gene="DKFP2p434C153"
polyA_site
2262
/gene="DKFP2p434C153"

ORIGIN
Query Match 45.4%; Score 174.6; DB 9; Length 2295;
Best Local Similarity 79.3%; Pred. No. 7e-30;
Matches 207; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 125 GCCCTGCCAGCCAGTTTCACAGAGGCTCTGAGGAATGAGAGGCGCGTGAAGGGCCACCA 184
Db 67 GCCCTGCCCTGCCAGTTCATGAGAGTATGAGAAACCAAGGCCACAGAAAGGGGTACA 126

QY 185 GCCATGTTGTGTGTAAGTCAAGTGGCCCTCTGAGAGTGGAGAAAGGGCCCGAG 244
Db 127 GTCACATTGCAATGTAAGCTGAGAAAGCGGCCCGCTGGAGTGGAGAAAGGGCCCAAC 186

QY 245 AACCTCAGATGGGACAGATACATCTGAGGAGGAGGAGGACCGAGTGTGAGCTGCAG 304
Db 187 ACCCTCAAAGATGGGGACAGGTACAGCCTGAAGCAGGATGGGACCGAGTTGTGAGCTGCAG 246

QY 305 ATCTGTGGCCCTGGCCATGGCGGAGCGCGGGAGTACTTGTGTGTCGGGCGAGAGG 364
Db 247 ATCTGTGGCCCTGGTCAATGAGAGTCTGAGAGTACTCTGTCATATGTGAGCAGAGAGG 306

QY 365 ACCTCAGCAGCTCACCATC 385
Db 307 ACCTCGCCAGCTCACTGTC 327

RESULT 13
BV014682 BV014682 621 bp DNA linear STS 30-MAY-2003
LOCUS S212P6028FPH4.T0 CZECHII/BI Mus musculus STS genomic, sequence
DEFINITION tagged site.
ACCESSION BV014682
VERSION BV014682.1 GI:31098577
KEYWORDS STS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Wade, C.M., Kulbokas, E.J. III, Kirby, A.W., Zody, M.C., Mullikin, J.C.,
Lander, E.S., Lindblad-Toh, K. and Daly, M.J.
JOURNAL The mosaic structure of variation in the laboratory mouse genome
MEDLINE Nature 420 (6915), 574-578 (2002)
PUBMED 22354684
COMMENT 12466852

Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersli@genome.wi.mit.edu
Primer A: None
Primer B: None
STS size: 621
Protocol:
WGS-discovery: Paired-end low-coverage whole genome shotgun reads
were generated from 129SI/SvimJ, C3H/HeJ, and BALB/cByg. The WGS
reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP

```



detection was carried out by SSAHA-SNP. 225,000 reads were annotated

as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J and the strain from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%.

## FEATURES

source  
1..621  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="CZECHII/Ei"  
/db\_xref="taxon:10090"  
/map="11 22-567 59585524-59586069"  
/clone\_lib="CZECHII/Ei"  
<1..621

## STS

## ORIGIN

Query Match 42.9%; Score 165; DB 11; Length 621;  
Best Local Similarity 77.0%; Pred. No. 1.4e-27;  
Matches 201; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
  
QY 125 GCCCTGCCAGCAATTCACAGAGGTTCTGAGGATGAGAGGCGCTGGGAAGGGGCCACA 184  
DB 477 GCCCTGCCCTAGATTATAGAGATTGAGAACCCAGAGGCCAGGGAGGTACTGTG 418  
QY 185 GCCATGTTGTGTGAATCAGCAAGGTGCCCTCTGTGAGTGGAGGAAGGGCCCGAG 244  
DB 417 GCCATCTGAGTCCAGATGAGCAAGGCTCCCCAGTGGAGTGGAGAGGGGTGAGAA 358  
QY 245 AACCTCAGATGGGACAGATATATCTCTGAGGAGGAGGGACACAGGTGTGAGCTCGAG 304  
DB 357 ACCCTGAAGATGGGGACAGATACAGCCCTGAGGAGGAGTGGGCGCATGTGTGAGCTCGAG 298  
QY 305 ATCTGTGCGCTGGCCATCGCGGAGCGCGGGAGTACTGTGTGTGCGGGCAGGAGGAGG 364  
DB 297 ATCCGTGACCTGGCTGTGGAGATCTGCGGAGTACTCACTGTGTGTGGCAGGAGAG 238  
QY 365 ACCTCAGCCAGCTCACCATC 385  
DB 237 ACCTCAGCCACACTGAGTGTCT 217

## RESULT 14

## AC139630

LOCUS Taxifugu rubripes clone 221D8, WORKING DRAFT SEQUENCE, 6 ordered pieces.  
DEFINITION AC139630 98613 bp DNA linear HTG 08-MAR-2003

## ACCESSION

AC139630 GI:28882132

## VERSION

HTG; HTGS\_PHASE2; HTGS\_DRAFT.

## KEYWORDS

Taxifugu rubripes (Fugu rubripes)

## SOURCE

Taxifugu rubripes

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Taxifugu.

## REFERENCE

1 (bases 1 to 98613)  
Akter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J., Hachichi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laic,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B., Magulies,E.H., Mastello,C., Masker,B., McDowell,J., Paquirigan,C., Pearson,R., Portnoy,M.E., Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.  
NISC Comparative Sequencing Initiative

## TITLE

## JOURNAL

## REFERENCE

## Green,E.D.

## Direct Submission

## TITLE

## JOURNAL

2 (bases 1 to 98613)

Green,E.D.

Direct Submission

Submitted (07-FEB-2003) NIH Intramural Sequencing Center, 8717

Groventom Circle, Gaithersburg, MD 20877, USA

## REFERENCE

3 (bases 1 to 98613)

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Green,E.D.  
Direct Submission  
Submitted (08-MAR-2003) NIH Intramural Sequencing Center, 8717  
Groventom Circle, Gaithersburg, MD 20877, USA  
On Mar 8, 2003 this sequence version replaced gi:28269380.  
----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov  
Contact: nisc.zoo@nhgri.nih.gov  
----- Project Information  
Center project name: eqw  
Center clone name: 221D08

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 97841 bases at least Q40  
Consensus quality: 98004 bases at least Q30  
Consensus quality: 98090 bases at least Q20  
Insert size: 85000; agarose-fp  
Insert size: 98113; sum-of-contigs  
Quality coverage: 11.5ix in Q20 bases; agarose-fp  
Quality coverage: 9.97x in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 28467: contig of 28467 bp in length

\* 28468 28567: gap of unknown length

\* 28568 31135: contig of 2568 bp in length

\* 31136 31235: gap of unknown length

\* 31236 41487: contig of 10252 bp in length

\* 41488 41587: gap of unknown length

\* 41588 74365: contig of 32778 bp in length

\* 74366 74465: gap of unknown length

\* 74466 96541: contig of 22076 bp in length

\* 96542 96641: gap of unknown length

\* 96642 98613: contig of 1972 bp in length.

## FEATURES

## source

1..98613  
/organism="Taxifugu rubripes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:31033"  
/clone="221D8"  
/clone\_lib="Incyte Genomics"  
1..28467  
/note="assembly\_fragment  
clone\_end:17  
vector\_side:left"  
28568..31135  
/note="assembly\_fragment"  
31236..41487

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

```
/note="assembly_fragment"
41588..74365
/note="assembly_fragment"
74466..96541
/note="assembly_fragment"
96642..98613
/note="assembly_fragment
clone end:SP6
vector_side:right"

ORIGIN
Query Match 18.5%; Score 71.4; DB 2; Length 98613;
Best Local Similarity 57.0%; Pred. No. 1.5e-06;
Matches 151; Conservative 0; Mismatches 111; Indels 3; Gaps 1;

QY 120 TGAAGGCCCTCCAGCCAGTTTCACAGAGGCTTCAGGAATGAAGAGGCCGCTGGAAGGGG 179
Db 61703 TGACAGGCTTACCAACCGACTTTCAGACGGCTCTCATCAACAGGAAGTCAATGAGAGGCA 61762
QY 180 CCACAGCCATCTGTGTGTGAACCTGAGCAAGTGGGCCCT---GTGGAGTGGAGGAAGG 236
Db 61763 ACAGCTGATCTGCAGCTGAGCTCAGCAAGCTGTTCTTCGGTGGAGTGGAGGAGAG 61822
QY 237 GGCCCGAGAACTCAGAGATGGGACAGATACATCCTGAGGCGAGAGGGACCAAGTGTG 296
Db 61823 GAGGAGAGCTCTAAAGCATGGAGACAAATATCAGTGTGAGGAAAAAAGAGCTGAGGTGG 61882
QY 297 AGCTGCAGATCTGTGGCTGGCCATGGCGACGCCGGGAGTACTTGTGTGTGGGGC 356
Db 61883 AGATGAGATCTCGACCTAGTGTGAGCGCGCGAGAGTACAGCTGTGTGTGGAG 61942
QY 357 AGGAGAGACCTCAGCCACGCTCAC 381
Db 61943 AACAGGCAACGAGAGCGCGTGATCAC 61967

RESULT 15
AX834822
LOCUS AX834822 2534 bp DNA linear PAT 15-DEC-2003
DEFINITION Sequence 1946 from Patent EP1347046.
ACCESSION AX834822
VERSION AX834822.1 GI:39920957
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,
Yanamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,
Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Naganari, K. and
Masuho, Y.
TITLE Full-length cDNA sequences
JOURNAL Patent: EP 1347046-A 1946 24-SEP-2003;
Research Association for Biotechnology (JP)
FEATURES
Source 1, 2534
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 18.2%; Score 70; DB 6; Length 2534;
Best Local Similarity 51.4%; Pred. No. 6e-06;
Matches 188; Conservative 0; Mismatches 175; Indels 3; Gaps 1;

QY 17 GACGGGGCTGTGCGAGCTGCAGATCTGTGGCTGCTGTGCAGATGTGGGGAGTAC 76
Db 937 GAGGGCAAGATGGCCATGCTGTGTATCCGGGGGCTCGCTCAAGGACCGCGCGGAGTAC 996
QY 77 TCCGTGTGTGTGGGGAGGAGGACCTCTGCCACTCTCACCGTGAAGGCCCTGCCAGCC 136
Db 997 ACGTGTGAGGTGGAGGCTTCCAGAGCACAGCCAGCCCTCCATGTGGAAGAAAGCAAC 1056
```

137 AAGTTCACAGAGGCTCTGAGGAATCAAGAGGCCGTGGAGGGGCCACAGCCATGTTGTGG 196  
1057 TGCCTTCACAGAGGAGCTGACCAATCTGCAGGTGGAGGAGAAAGGCACAGCTGTGTTACG 1116  
197 TGTGA---ACTGAGCAAGGTGGCCCTCTGTGAGTGGAGAAAGGGGCCGAGAACTCAGA 253  
1117 TCAAGACGGAGCACCCCGCGGCCACAGTGACCTGGCGCAAGGGCCTCTTGGAGTACGG 1176  
254 GATGGGACAGATACATCTCTGAGGAGGAGGGGACCAGGTGTGAGTGCAGATCTGTGGC 313  
1177 GCCTCAGGGAAGCACACAGCCAGCCAGAGGGGCCCTGACCCCTGCGGCTCACCATCAGTGCC 1236  
314 CTGGCCATGGCGGAGCGCGGGAGTACTTGTGTGTGTCGGGGCAGGAGACCTCAGCC 373  
1237 CTGGAGAAGGACAGACGACACCTATACCTTGCACATTTGSCCAGGCCCATGTCGGGGCC 1296  
374 ACGCTC 379  
1297 CAGCTC 1302

Search completed: September 19, 2004, 19:27:07  
Job time : 1574.62 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 19, 2004, 14:16:08 ; Search time 4363.43 Seconds  
(without alignments)  
10628.572 Million cell updates/sec

Title: US-10-077-130-4\_COPY\_1\_1070

Perfect score: 1070

Sequence: 1 tgcctaccagagccacac.....tagtgcgagccgcgggtt 1070

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_on.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_nu.\*  
20: em\_on.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1070	100.0	20435	9	HSJ2535	AJ002535 Homo sapi
2	1005	93.9	95745	9	AL359510	AL359510 Human DNA
3	720	67.3	211829	10	AL645854	AL645854 Mouse DNA
4	699.2	65.3	260938	2	AC099089	AC099089 Rattus no
5	527.6	49.3	44358	2	AC142478	AC142478 Rattus no
6	505	47.2	7531	9	HSJ2535	AJ002535 Homo sapi
7	291.4	27.2	1198	11	MMU409657	MMU409657 Mus muscu
8	255	23.8	2884	6	AX714317	AX714317 Sequence
9	255	23.8	2884	9	AK056556	AK056556 Homo sapi
10	199.2	18.6	3316	9	BC007201	BC007201 Homo sapi
11	197.8	18.5	135033	9	AC009955	AC009955 Homo sapi
12	197.8	18.5	186883	2	AC040991	AC040991 Homo sapi
13	181	16.9	225471	2	AC112361	AC112361 Rattus no
14	181	16.9	266167	2	AC121633	AC121633 Rattus no
15	166.8	15.6	187999	2	AC115011	AC115011 Mus muscu
16	158.6	14.8	2141	10	SS6581	SS6581 alpha inhib
17	126	11.8	5382	9	AB014557	AB014557 Homo sapi
18	117.2	11.0	471	6	BD060356	BD060356 Secreted
19	108.8	10.2	153945	2	BX322540	BX322540 Danio rer
20	99.6	9.3	82746	1	AF453501	AF453501 Actinosyn
21	98	9.2	135964	9	AL353893	AL353893 Human DNA
22	95.6	8.9	745	9	HSJ2535	AJ002535 Homo sapi
23	90.6	8.5	1393	11	PM11H12G	PM11H12G
24	88.4	8.3	154890	2	BX548249	BX548249 Danio rer
25	84.8	7.9	1065	11	PM2B12B	PM2B12B
26	84.6	7.9	1393	11	PM11H12G	PM11H12G
27	84.4	7.9	1279	11	PM2H12G	PM2H12G
28	83.6	7.8	95209	2	AP004323	AP004323 Oryza sat
29	82.2	7.7	181850	2	BX276102	BX276102 Danio rer
30	81	7.6	956	11	PM2D12B	PM2D12B
31	79.8	7.5	976	11	PM12B6G	PM12B6G
32	79.8	7.5	1065	11	PM2B12B	PM2B12B
33	79.6	7.4	956	11	PM2D12B	PM2D12B
34	79.4	7.4	733	9	HSJ2535	AJ002535 Homo sapi
35	79.2	7.4	301925	1	AP005046	AP005046 Streptomy
36	79	7.4	125020	9	AF429315	AF429315 Homo sapi
37	78.6	7.3	10444	14	PVI422133	PVI422133 Pseudorab
38	78.4	7.3	224777	2	AC138109	AC138109 Mus muscu
39	78.2	7.3	11059	14	PRVULGNS	PRVULGNS
40	77.8	7.3	869	11	PM2A12B	PM2A12B
41	77.4	7.2	1052	11	PM2H12B	PM2H12B
42	77.4	7.2	162832	5	BX005391	BX005391 Zebrafish
43	77.4	7.2	224777	2	AC138109	AC138109 Mus muscu
44	77.2	7.2	991	11	PM12H12B	PM12H12B
45	77.2	7.2	1143	11	PM12A11G	PM12A11G

ALIGNMENTS

RESULT 1  
HSAJ2535  
LOCUS  
Homo sapiens mRNA for obscurin (OBSN gene).  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

20435 bp mRNA linear PRI 14-SEP-2001

GI:15026973

OBSN gene; obscurin.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

Young, P., Ehler, E. and Gaucel, M.

Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly

J. Cell Biol. 154 (1), 123-136 (2001)  
 21342081  
 MEDLINE  
 PUBMED  
 11448995  
 2 (bases 1 to 20435)  
 REFERENCE  
 AUTHORS  
 Gautel, M.S.  
 TITLE  
 Direct Submission  
 Submitted (31-OCT-1997) Gautel M.S., Structural Biology Division,  
 European Molecular Biology Laboratory, Meierhofstr. 1, Heidelberg,  
 69117, GERMANY  
 JOURNAL  
 Revised by author 20-JUL-2001  
 REMARK  
 FEATURES  
 Location/Qualifiers  
 1. 20435  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="1q42.1"  
 /cell\_type="cardiomyocyte"  
 /tissue\_type="cardiac"  
 /dev\_stage="adult"  
 1. 20435  
 /gene="OBSCN"  
 72. .19934  
 /gene="OBSCN"  
 /function="titin binding"  
 /codon\_start=1  
 /evidence=experimental  
 /product="obscurin"  
 /protein\_id="CAC44768.1"  
 /db\_xref="GI:15026974"  
 /db\_xref="GOA:Q96AA2"  
 /db\_xref="SPRTEMBL:Q96AA2"  
 /translation="MDQPFSGAPRFLTPKAFVSVSGKDATLSQIVGNPQVSWSE  
 KDQPVTAGARFLAQDGLYRLTILDALGDSQVVCARNAGFAFAVAGVLDVDAE  
 AACAEQAPFLHPTSLRVREGSEATFRCHVGSGSPAPVSWKDGRLPEGPGRVVR  
 ELGEASALIRAIRAPRDGGTYVRAENPLGASAAALVDSADADTSRGTSTAA  
 LLAHLQRRAARAGPAPSPSTGRTCTCTCTGKHARLSVYGTGPKPTVTKDQDQ  
 LVTEGRHHVYDAQENFLIKFCQSDRGLYTCASNLVGQTSVSLVVRREPAPV  
 FKRLQDLVREKESATFLCEVPQSTEARAFKEETRLWASAKYGTIEEGTETLTVR  
 NVSADNOEVVICTPEGSTVAELAVQGNLLRLPKRTAVRVDGTAMFCVELAVPGP  
 VHLNRDAEYVAGGRVAISAEGRHTLTISQCCLEDVGOVAFMAGDCQSTSTPCVSP  
 RKPLPDPVDPVAKMESVILSWSPPHGEPPTIDGVLEKXKLGYTWIRCHEA  
 EWATPELTVDVAENSGFVRVSNLNSFGQSPYLEFPGTVHLKPLAVRTLKAVQA  
 VEGEVFTVLDLTVAENSGFVLDQALKASSVYIEHCRTHTLTIREVPASDAQAL  
 KFNANGTESIRMEVRAAPGLTANKPPAAAREVLAHLHEEAQGLTAELEDQAAATVL  
 KDGLTSPGPKYEVQASGRVLLVRDVARDDAGLYECVSRGRTAYQLSVQGLAFLL  
 HKMAGSCDVAAGGPAQPCETSEAHVHVHWYDKMGLCHSGERFLOEDVGTTHRLV  
 AATVTRDEGTSCVCGEDSDVDFRLVSEPKVVFVFAKEQLARRKLOAEAGSATLSCEV  
 AQTQETVWKDGKKLSKSSKVCMEATGCTRLVAVQAGQADAGEVSCAGGRLSPH  
 LDVKEPVFAKQVQAHSEVQAEAGANATLSCEVAQAQAEVWYKDKGKLSKSLKHV  
 EAGCRRLVYQAGKTDAGDYCEARGQKVSFLRHLITPEKMFVFAKEQSVHNEVQAE  
 GAGAMLSCEVAQATETVTKDGKLSKSSKVGMEVKGCTRLVLPQAGKADAGEYSC  
 EAGQRVSPFLHITPEKGVFAKEQSVHNEVQAEAGTAMLSCEVAQATETVTKDGK  
 KLSKSSKVRMEVKGCTRLVVOQKADAGEVCEAGGQVSPQLHITPEKAVFAKEQ  
 LVHNEVTEAGSATLSCEVAQATETVTKDGKLSKSSKVRIBAAQCMQRLVVOQA  
 QADAGQVCEAGGRLSPHLDVSEPKAVFAKEQLARRKLOAEAGSATLSCEVAQAO  
 TETVTKDGKLSKSSKVRMEAVGCTRLVVOQAQADTGEVSCAGGRLSPHLDVA  
 EPKVFAKEQVHREVQAGASTTSCVEVAQATETVWYKDKGKLSKSSKVRIBAAQ  
 CTRRLVVOQAQADAGEVCEAGGRLSPHLDVSEPKAVFAKEQVQAEVQAEAGTAMLS  
 TSCVEVAQATETVTKDGKLSKSSKVRMEAVGCTRLVVOQAQADAGEVSCAGD  
 LRLVFLHVAEPKVFVFAKEQVHREVQAGASTTSCVEVAQATETVTKDGKLSKSS  
 SKVRVAVAGCTRLVVOQAQADAGEVCEAGGRLSPHLDVSEPKAVFAKEQVQAE  
 PLVVKHEDILITATLPSAATVWLKQGVETRRSKRHTASQDHTLTLVHGAQVL  
 DSALYSCRVAEQDTPQVVEVAAKFCILLEPVCGLGTTVLACELSPACAEVVR  
 CGNTYRCRPFQVHREVQAGASTTSCVEVAQATETVWYKDKGKLSKSSKVRIBAAQ  
 FMGLSTVVAEEGEATFCQVSPSDVAVVWFDGALLQPSKFAISQSGAGSLTIS  
 DMLSDAGCTVVAEEGASAAALVREAPVLFKKLQPTVEERSSVLESTLTPW  
 ELRWTRNATALAPKQVTHEAGARHLVHNVGADRGGFCGTPDDTKLQAKLTVM  
 QRVLRLHVAEPKVFVFAKEQVHREVQAGASTTSCVEVAQATETVTKDGKLSKSS  
 LTLGLRPEDSGVLMVFAKEGHTVARSBLVTLVPSVSRPLQDVTVEKVTLECLLS  
 RPNVDVRLKDGVELGAPKTMATAAQACRSLTLYBCEPDQGVVCDHDAQSSASV  
 KVQGRYTLIYRVLADAGEIQVFAENAESRAQLRVKELPTLVRPLDKLAMEKHR  
 GVLEQCVSRASAAQVRWFKGQELQPGKYLSDGGLYRKLITSDVHAEDETITCDAG

DNKTSAQFFVEEQSITIVRGLQDVTVMPEPAPAFECETSIPSVRPPKWLKTKVLQAG  
 GNVGLEQGTVHRLMLRRTCTMTGPVHFTVGRKSRARLVSDIPVVLIRPLPKFTG  
 RELQSVLSCDFRPAKAVOVYKODTFLPSPEKFKMLEGOMAEELRLMLPADAGY  
 RCQAGSHSSTEVTEAREVTVTQDABATEEGWASFCSELSEDEEVEWLSNGVP  
 LYNSDFHEISHKRRHTLVLSIQADADAGIVRASSLVKVSARLEVRVVPVFKALD  
 DLSAEERTLALQCEVSDPEAHVVRKDGVLGSPDKYDPLHTAGTGLVVDVSPD  
 AGLTVCHVSEETARVVRHDLHVLTIKRLKTMELVEGSCSECVLSHESASDPAW  
 TVGKTVGSSRFQATQGRKYLIVVREAPSDAGEVVFVRGLTSVSLRVRPAA  
 IIKPLEDQWAPGEDVELRCELSRAGTPVHMLKDKRKAKIRSKQYKVVCEGMLVLR  
 CASLKDAGEYCEVASKSTASLHVEKANCFTBELTNLQVEEKGTAFTCTKEHPAA  
 TWTWRKGLLELRASGKQPSQEGTLRLTILSALEKADSDTVCTIGQAGRAQLVQ  
 RVHIIEDLVDOEGSSATFCRIRIPANVEPVHFLDKTPLHNLNIDIAOPGGY  
 HVLRLALAKDSGTIVFEAGDORASALVRTEKPSVFSRELTDATITEGEDTLVCE  
 TSTCDIPMCWTQDKTLRGSARCOLSHGHRQALLITGATLQDSGRYKSCAGGACSS  
 IVRVARPVREAEALKOLEVLEGGANILRCSLVSSVAAPVKVCYGNVVRFGDKYLSKQ  
 EGAMLELVNRLPDSDGRYSCFGDQTTTATLTVTALPAQFICKLNRKKEGATGAT  
 LRCELSKTAPVWRKSETLRDGRYCLRDQDGMCELOIRGLANVDAEYSCVCGEER  
 TSASLTIRPMAHFIRGLRHOESTEGATLRLCELSKAAPVWRKGRSLDGRHSL  
 RODGAVCELOICGLAVADAGEYSCVCGEERTSATLVKALPAKFTGLEGRNEAVEGAT  
 AMLWCELSKAPVWRKGPENLRDGRYILRQEGTRCELOICGLAMADAGEYLCVCGQ  
 ERTSATLIRALPARFIEDVKNQAREGATVLOCELSAAPVWRKGRSLRDRDY  
 SLRQGTICELQIRGLAVADTGEYSCVCGERTSAMLTVALFIKFTGELRNEATEG  
 ATAVLRCELSKAAPVWRKGRSLRQDGRHSLRQDARGCELOIRGLVADAGEYLCMC  
 KERTSAMLTVRAMPKFI EGLRNEATEGDTATLWCELSKAAPVWRKGRHETLRDGD  
 RHLRQDSRCELOIRGLAVADAGEYSCVCGERTSATLVRALPARFIEDVKNQEAR  
 EGATAVLOCELSKAAPVWRKGRSETLRDGRYSLRQDTRCELOICHLGVADTGEYSC  
 VCGOERTSATLVRAPQVPREPLQSLQAEBSGTATLQCELSBPTAVVASKGLOLQ  
 ANGRERPLQGTAEVLQDLQREDTGEYTCGSOATSLTVAAPVRLRLELOHQ  
 BVDEGTHALCCELSRAGAVWRKGRSLQIPCKAYQVQDGAALVLRGQEDAG  
 DYTCDTGHQSMASLVVRPPKFLRQSLQEGTDIARLCCOLSDAESGAVVQMLK  
 EGVELHAGPYKEMRSQATRELILHQLQEDTGEYACVCTGQKTAASLRVTEPEVTV  
 RGLVDAEYTDDEVEFCEVSRAGATGVQMLQPLQSNVEVAVRDRGRIHRLK  
 GVTPEDAGTVFHLGNEHASAQLTVRAPEVITILEPLQDVLQSGQDASFCQSLRASG  
 QEARWALGVLOANEMNDITVEOGTLLHLLTKVTLLEDAGTVSFHGTCSSEAQLK  
 TAKTVYKGLNLEVALRGGALFECOLSQPEVAHTWLLDDEPVRTSENAEVPFENG  
 LRHLLLLKLNLPQDSCTVTLAGDMVTSATLVGRWLELELEPLKXNAVAGAQART  
 LTSEAVFVSGASWYINGAAVQDDSDMTVADGSHQALLRLEPLKXNAVAGAQART  
 AVASARLVLPDPPDEDAVVAHSHSTVLSWAAPMDSGGGLGVRYRVEKAGATQ  
 WRLCHELVGPCVVDGLPAGETVFRVAAVPGVAGEPVHLPQTLVRLAPFPVPPQ  
 PSAPESQVAAAGEDVLELVAEAGEVIMWKGEMRIQPGREFVVSQSGQMLVWKG  
 FTADQGEVHCLGAQSGICPAATFOVALSPVDEAPQSLPPEAAQEGDLHLVIEA  
 LARVMSREPTLSISELPEEDGSRORLOAEAEVAPDLSEGVSTADELARLGDADL  
 SHTSDEDSRAGTSLVLYLKKAGRPCTSLASKVGAAPAPSEGVSPQQQPFLAARVP  
 LGDLSLTDGSPMDKAAVYLKQAFKGVKVKENKQEGFMSFTFGDTPAEVADAR  
 LECVYAKDAVRAWMLKDVGLDGRHHHLDQDGGTCSLLIAGLRDAGAGTCQVS  
 NKFCQVTHSACVVSSESAESSGGLDDAFRAARLRLHFRTRSPAEVDEEFL  
 LSADGPAEPADWOTYREDEHFCIRFEALTTEARQAVTRFOEMFATIGVIEKL  
 VEOGRPVMECISKETPAVVPPEPLSLTSDAAPVELTELQNOEVDQGVPPVDFCV  
 VTGQMPSVWFKDKLLEDDHVMINEDQGGHLLITAVPADMGVYRCLAENSMG  
 VSTKABLRDLTSDTDADATESSSYSAQYLSRREGTSTDSGQLPQVVE  
 ELRDLQVAPGRFLAKFQKVKRGPAPRYWFKQGPATASAHRTMDTKLILHLEITIS  
 VTRDSQGYAAYISNANGAAYSSARLVGRDPEEPKPSADVHEQLVPPMLRFPPTK  
 KVKKSGTITTSVKVEPVPVTHWLRREEARGVLMIGPTPGTVASSAQHSLJLD  
 VROHQGTYCTIASNAGAOALCSALHVSGLPKVEOEKVEKALISFTFLOGTQALSA  
 QGQTFADLGGORKEEPLAAKALHSLAEVGTETFLQKLTSTQTEMVSAKITQA  
 KLOYPQGSDEDSKTPSASPRHGRSPSSSQESSSESDGARGEIFDLYVVTADLY  
 PLGAQAQAILIREQYVEVLDAAHPRLWRTPTKSPSRQGVWSAYLDRKLSP  
 EWGAAPAEFPEGEAVSEDEYKALSVIQELLSQAFVEELQSHHQLHQRCHP  
 VPIAVAQKAVIFRNRDIGHSSFLQELQCCDDDDVAMCFKQAAFEVQLFELV  
 GRVQAEVYVSTAIETFEYKYAEALQADPSQPPPPQLHYLEQVRYQVQALLK  
 ELIENKARNQNCALLQAVVAVSALPQRAENKLSHLMENYFOTLQALGEPHQHF  
 IVMGAPKQNKGNHNVLPFNHLVICRPRDSDTDTVSVFNNMLKSLIDLND  
 QVEDDRAFEVWQEREDSVRYLLQARTLIIKSSWKEICGIIQORLALPYRPPDPE  
 ELACTAELGETVKLACRVGTPTKPYLWYKDKGKAVQVDPHHILIEDPDGSCALIDS  
 LTVGDSQKQCFMFAASAGNSTGLKILQVFPFRVKNLASPFVEGEDAQTCTIEGA  
 PYPQIRVQKGLITTKNFQTSLEPSRGLLVLRRAKEDJGLVECELVNRKISAR  
 ASALRILQSMLOAQEQCHEEQLVAEVEVTEKVPKKTVIIETITTVTKSPQGR  
 RSPKSPSRSPSRCSAPLRLGLADLLYLPAGQPRRPAEPGQVVPVPTLYTEA  
 EAHSPALPQSGQPKWVEVEETIEVRKMGQGVSTPEVRSKSHGLFTLPAGATP  
 GGDNSNNKLLAQAEAAQGTAVRVDVARGSVDMWASGMSLEBEGTM  
 BEAGEEREGDGFATVEESODTHSLGDRDPKILTHNGMLTLADLSDYVPEGETPHC  
 GPGPGAPDPPCEVSVIREIGEPTVQGVLLSVGHALGPRGLGFLPREPGASPP  
 GPGVRSLEGSFLLREAPAPVSGAPWTSQFCTIRRSADSGSSFTTEISTQTVNPG

TVGETVTLHICPDRDGEAAOP"

```
Query Match      100.0%; Score 1070; DB 9; Length 20435;
Best Local Similarity 100.0%; Pred. No. 1.3e-119;
Matches 1070; Conservative 0; Mismatches 0; Indels 0;
```

QY	1	TGCGTACAGAGGCCCA	CAC	TCCGGCCGCTG	CCAGAGCCCCCA	TATAGAGAGAGGT	TCCCGC	60
DB	1	TGCGTACAGAGGCCCA	CAC	TCCGGCCGCTG	CCAGAGCCCCCA	TATAGAGAGAGGT	TCCCGC	60
QY	61	CCGCCACCGTCATGGAT	CAG	CCACAGTTC	CAGCGGGGCGCCCG	CTTTCTTCA	CCCGGCCCCA	120
DB	61	CCGCCACCGTCATGGAT	CAG	CCACAGTTC	CAGCGGGGCGCCCG	CTTTCTTCA	CCCGGCCCCA	120
QY	121	AGGCTTCGTGTGTGCGGT	GGGCAAG	CAGCCAC	ACCTTCAG	TGTGCCAGATCGTGGGTAA	TTC	180
DB	121	AGGCTTCGTGTGTGCGGT	GGGCAAG	CAGCCAC	ACCTTCAG	TGTGCCAGATCGTGGGTAA	TTC	180
QY	181	CCAGCCACAGGTGAGCT	TGGGAGAAG	CACAGACCG	GTGACGGCGCGCGCGCTTCC			240
DB	181	CCAGCCACAGGTGAGCT	TGGGAGAAG	CACAGACCG	GTGACGGCGCGCGCGCTTCC			240
QY	241	GTGTGGCCAGGACGGCGA	CCCTTAC	CGCCTCA	TATCTTGAGAC	CTGCGCTGGCGGACA		300
DB	241	GTGTGGCCAGGACGGCGA	CCCTTAC	CGCCTCA	TATCTTGAGAC	CTGCGCTGGCGGACA		300
QY	301	GTGGGCAATACGTGTG	CCGCGCGCA	TATGCC	TATAGCGAGGCTTCG	CTCCGTGGGCTC		360
DB	301	GTGGGCAATACGTGTG	CCGCGCGCA	TATGCC	TATAGCGAGGCTTCG	CTCCGTGGGCTC		360
QY	361	TGCAGGTGGACCGGGA	GCGCGTGC	CGCCAG	CAGGCGCGCA	TTCTCTGTGCGGCCCA		420
DB	361	TGCAGGTGGACCGGGA	GCGCGTGC	CGCCAG	CAGGCGCGCA	TTCTCTGTGCGGCCCA		420
QY	421	CGTCCATCCCGGTGCG	CAGGGCTC	AGAGGCA	CACTTCCG	TGTGCGCGTGGGTGGCTTCC		480
DB	421	CGTCCATCCCGGTGCG	CAGGGCTC	AGAGGCA	CACTTCCG	TGTGCGCGTGGGTGGCTTCC		480
QY	481	CGAGGCGGCGAGTGA	GCTGTG	CAAGAC	CGGGCGGCTTGGG	TGAGCCCGACGGGCCCC		540
DB	481	CGAGGCGGCGAGTGA	GCTGTG	CAAGAC	CGGGCGGCTTGGG	TGAGCCCGACGGGCCCC		540
QY	541	GCGTGCGGTGGAGGAG	CTCGGCGAG	CAAGTGC	CGCTGCGCATTCGG	CGCGCGCGCGGCCG		600
DB	541	GCGTGCGGTGGAGGAG	CTCGGCGAG	CAAGTGC	CGCTGCGCATTCGG	CGCGCGCGCGGCCG		600
QY	601	GCGACGCGGCAC	CTTAC	AGAGTCCG	CGCGAGAC	CCGCTGGCGCTCCAGCGCCGCG		660
DB	601	GCGACGCGGCAC	CTTAC	AGAGTCCG	CGCGAGAC	CCGCTGGCGCTCCAGCGCCGCG		660
QY	661	CGGCGCTAGTGTGGA	CTCGGAC	CGCCCGGACA	CGGCGACCGCGCCCGGGAC	CTTCA	CGG	720
DB	661	CGGCGCTAGTGTGGA	CTCGGAC	CGCCCGGACA	CGGCGACCGCGCCCGGGAC	CTTCA	CGG	720
QY	721	CCGGCTCTTGGCGCA	CTTGA	CGGGCGGCG	CGAGGCTATGCG	CGCGCGAGGGCGCCCCCG		780
DB	721	CCGGCTCTTGGCGCA	CTTGA	CGGGCGGCG	CGAGGCTATGCG	CGCGCGAGGGCGCCCCCG		780
QY	781	CCCTCACCGCCAG	CACCGGCA	CGGCA	CCCTG	CAGGCAAGCAGCGCGGCC		840
DB	781	CCCTCACCGCCAG	CACCGGCA	CGGCA	CCCTG	CAGGCAAGCAGCGCGGCC		840
QY	841	TCAGCTGCTACGTGA	CCGGCGAG	CCCCAAG	CGCCGAGACGGTGTG	GGAAGAGACGGCCAGC		900
DB	841	TCAGCTGCTACGTGA	CCGGCGAG	CCCCAAG	CGCCGAGACGGTGTG	GGAAGAGACGGCCAGC		900
QY	901	TGGTGACGAGGGCG	CGCGCAC	CTGTGTG	TACAGGAGCG	CGACGAGAACTTCGTGCTCA		960
DB	901	TGGTGACGAGGGCG	CGCGCAC	CTGTGTG	TACAGGAGCG	CGACGAGAACTTCGTGCTCA		960
QY	961	AGATCTCTTCTG	CAAGCAG	TGCGAC	CGCGGCTCT	TACACTGTCACGGCGCTCA	ACCTCG	1020

Dd 961 AGATCTCTTTCGACAGCAGCTCGACGCCGGCCCTCTACACTGCAGCGGCTCGCACTCG 102

Qy 1021 TGGCCAGACCTACAGCTCTGTGCTGTGCTAGTGCGGAGCCCGGGTT 1070  
1021 TGGCCAGACCTACAGCTCTGTGCTGTGCTAGTGCGGAGCCCGGGTT 1070

Dd 1021 TGGCCAGACCTACAGCTCTGTGCTGTGCTAGTGCGGAGCCCGGGTT 1070

RESULT 2

AL359510 Human DNA 95745 bp linear PRI 02-FEB-2002

LOCUS Human DNA sequence from clone RP11-520H14 on chromosome 1, complete sequence.

DEFINITION

ACCESSION AL359510

VERSION AL359510

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 95745)

AUTHORS Blakely,S.

TITLE Direct Submission

JOURNAL Submitted (02-FEB-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humkey@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Feb 4, 2002 this sequence version replaced gi:18121469. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chrl RP11-520H14 is from the library RPI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBACe3.6

FEATURES

source Location/Qualifiers

1..95745

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="1"

/clone="RP11-520H14"

/clone\_lib="RPI-11.2"

20318..20426

/note="Sequence from uni-directional dGFP big dye terminator reads only."

93348..93408

/misc\_feature

/note="Single clone region. Assembly confirmed by restriction digest data."

ORIGIN



```

119561 CGCGCTTCCGCTGCGCCAGGACGGGATGTATACCGCTCACAATTTCTGATCTGCGCG 119502
292 TGGCGGACAGTGGCAATAGTGTGCGCGCGGCAATGCAATAGGCGGCGCTTGGCTG 351
119501 TAGTGACAGGCGGACGATGTGTGTCGCGGAGGAGACGCATAGGGAGGCGCTTGGCG 119442
352 CCGTGGGCGCTGACGAGTGGAGCGGAGCGCGCTGCGCGGAGGCGGCGCGCACTTCTGTC 411
119441 CCGTGGGCGCTGACGAGTGGAGCGGAGCGCGCTGCGCGGAGGCGGCGCGCACTTCTGTC 119382
412 TGGCGGCGGCTGACGAGTGGAGCGGAGCGCGCTGCGCGGAGGCGGCGCGCACTTCTGTC 471
119381 TGGCGGCGGCTGACGAGTGGAGCGGAGCGCGCTGCGCGGAGGCGGCGCGCACTTCTGTC 119322
472 GTGCGCTTCCGAGCGCGGAGTGGAGCGGAGCGCGCTGCGCGGAGGCGGCGCGCACTTCTGTC 531
119321 GAGGCTCACCAACCTGCTGTGAGCTGTTCCAAAGATGGCGGCGCTTGGCGCGGCGG 119262
532 ACGGCGCGGCTGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGG 591
119261 ATGCGCGCGGCTGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGG 119202
592 CGCGCGCGGCTGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 651
119201 CAAGGCTCGGATGTTGGACCTACGAGTCCGAGCAGAAATCCGCTGGGCTTGGCGCA 119142
652 CGCGCGCGGCTGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 711
119141 CGCGTGGCGGCTGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 119085
712 CCGCGCGGCTGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 771
119084 CCGCGCGGCTGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 119025
772 CGCGCGCGGCTGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 831
119024 GCATACCGGCTTCTCCAGCGGCTGGACGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 118965
832 ACGCGCGGCTGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 891
118964 ACGCGCGGCTGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 118905
892 ACGCGCGGCTGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 951
118904 ACGCGCGGCTGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 118845
952 TCGTGCTCAAGTCTTCTGCAAGCAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 1011
118844 TCGTGCTCAAGTCTTCTGCAAGCAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 118785
1012 CCAAGCTGCGGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 1059
118784 CCAAGCTGCGGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAGTGGAGCGGAG 118737

```

```

RESULT 4
AC099089/c
LOCUS AC099089 260998 bp DNA linear HTG 10-MAY-2003
DEFINITION Rattus norvegicus clone CH230-154E3, WORKING DRAFT SEQUENCE.
ACCESSION AC099089
VERSION AC099089.5 GI:30522701
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Rattus.
1 (bases 1 to 260998)
Muzny,D,Marie., Metzker,M,lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,C., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Dugan,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Greggordis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunarane,P., Haaland,W., Hamil,C., Hamilton,C., Hernandez,J.,
Harvey,P., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Herrandez,R., Hines,S., Hladun,S.D., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowls,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Louised,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwakemelen,O., Okwuonu,G., Olarunpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sned,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soosa,D.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villamas,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 260998)
Worley,K.C.
Direct Submission
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 260998)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center

```

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GGLZ

Center clone name: CH230-154E3

----- Summary Statistics

Assembly program: Atlas 3.0

Consensus quality: 21829 bases at least Q40

Consensus quality: 21828 bases at least Q30

Consensus quality: 220845 bases at least Q20

Estimated insert size: 226118; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: this is a 'working draft' sequence. It currently

consists of 1 contigs. Gaps between the contigs

are represented as runs of N. The order of the pieces

is believed to be correct as given, however the sizes

of the gaps between them are based on estimates that have

provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 260998: contig of 260998 bp in length.

Location/Qualifiers

1. 260998

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-154E3"

1. 1486

/note="wgs\_contig"

## FEATURES

source

misc\_feature

## ORIGIN

```

Query Match      65.3%; Score 699.2; DB 2; Length 260998;
Best local similarity 81.5%; Pred. No. 1.3e-75;
Matches 822; Conservative 0; Mismatches 183; Indels 3; Gaps 1;

Qy 52 AGGTCCCGCGCCGACCGTTCATGGATCAGCCACACAGTTTCAGCGGGGCGCCCGCTTTCTCA 111
Db 246518 AGTCCCCCAACCCCTGTAAATCATCATGACCACTCTTCACGCGAGCACCCCGCTTCTGA 246459

Qy 112 CCGCGCCCAAGCCCTTCGTGTGTCGTGGCGAAGGACCCCTCAGTTCAGTCGATCG 171
Db 246458 CGCGGCCAAAGGCTTTTGTGTATCTGTGGCAAGGATGCCAGCTGAGCTGCCAGATCG 246399

Qy 172 TGGGTAATCCCAACCCACAGTGTAGCTGGGAGAGGACCAAGCAGCGGTGACGCGCGCG 231
Db 246398 TGGGCAACCCCAACCCACAGTGTAGCTGGGAGAGGACCCGAGCAGCTGGAGCGAGAG 246339

Qy 232 CGCGCTTCGTGTGGCCAGACGCGACCTTACCGCCCTCATCTATCTGGAGCTGGCGC 291
Db 246338 CACGCTTCGCGCTGGCCAGGAGGATGTATGACCGCTTACCACTCTCGATCTGGCTC 246279

Qy 292 TGGCGACAGTGGGCAATAGTGTGCGCGCGCGCAATGCATAGCGAGGCTTCGCTG 351
Db 246278 TAGGTGACAGGGGCACTAGTGTGTACGAGGAGGACCAATAGGGAGGCTTCGCGC 246219

Qy 352 CCGTGGCGCTTCAGTGTGAGCGCGAGCGCCGCTGCGCGAGCGCCCGCCTTCCTCGC 411
Db 246218 CTGTAGGTTTTCGAGTGTGACTCGAGGAGGACGCTGTGCGAGGAGGCGCCCTTCTCGC 246159

Qy 412 TGGCGCCCACTCCATCCGCGTGTGCGAGGCGTTCAGAGGCCACTTCCGTCGCGGTG 471
Db 246158 TGGCGCCCACTCCATCTGCTGTGCGAGGCGCGCAGCGCCACTTCCGATGTCGCTCG 246099

Qy 472 GTGCTCCCCAGCGCGCAGTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 531
Db 246098 CGGCTCACCGCAACCTGCTGTGAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 246039

```

```

Qy 532 ACGGCCCCCGCTGCGCTGTGGAGAGCTCGCGAGGCAAGTGCCTGCGCATTCGGGCGG 591
Db 246038 ATGCCCCCCCACGTGCGCTGTGAAGATCGCGAGAGGCGCGCTTCGCATTCGGTTCG 245979

Qy 592 CCGCGCCCGCGGACGCGCGGCACTTACGAGGTCGCGCGCGAGAACCCGCTGGGCGCTGCCA 651
Db 245978 CAAGGCTTCGCGATGCTGGCACTTACGAAATTCGAGCAGAGAAACCCACTGGGCTCCGCA 245919

Qy 652 GCGCGCGCGCGCGCTAGTGTGACTCGGAGCGCGGACACGCGCGAGCGCGCGCGCGGA 711
Db 245918 GCGCTGCGCGCGCTCTCGTGTGACTCGGATGCGGAGG---CTGCGGACACACCGGAA 245862

Qy 712 COTCCACGCGCGCGCTCTCGTGTGCGACCTGTGACGCGCGCGGAGGCTATGCGCGCGGAG 771
Db 245861 COTCGGTTGCCAGCTCTCTGCGCACCTGTGACGAGCGCGCGGCGGCGCATGCGCGCAGAG 245802

Qy 772 GCGCGCGCGCTCACCGCGCGGACCGGACGCGGACCTGCGAGCTGCGAGTGAAGCGAAGC 831
Db 245801 GCGTCCCTCTCTTCTCCACCTGCTGTCGCGCGGAGCCCGGAGCGGAGGCGAAGCGAAGC 245742

Qy 832 ACGCGCGCTCAGCTGCTTACGTGACCGCGGAGCGCGGAGCGCGGAGCGGAGGAGG 891
Db 245741 ACGCGCGCTCAGCTGCTTCTGACCGCGGAGCGCGGAGCGCGGAGCGGAGGAGG 245682

Qy 892 ACGCGCGCTGTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTACACCTGCGAGGAACT 951
Db 245681 ACGGCGAGCTGCTGAACGAGGCGCGGAGACACAGTATATATAGGAGTGTGAGGAGGAACT 245622

Qy 952 TCGTCTCAAGATCTCTTCTGCAAGCAGTCCGAGCGCGCGCGCTTACACCTGCGAGGCGT 1011
Db 245621 TCGTCTTAAATCTTTTGTGCAAGCAGTCTGATCGCGCGCTTACACCTGCGAGGAGT 245562

Qy 1012 CCAACCTGTCGGCGAGACCTTACAGCTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 1059
Db 245561 CCAACCTGTCGGCGAGACCTTACAGCTCGGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 245514

```

## RESULT 5

AC142478/c

LOCUS

DEFINITION

AC142478

AC142478.1 GI:29374186

HTG; HTGS\_PHASE1

Rattus norvegicus (Norway rat)

SOURCE

ORGANISM

Rattus

REFERENCE

AUTHORS

Murny D. Maris., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila, M.B., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Georgopoulos, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guavara, W.,

Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hughes, M.,

Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,





QY 412 TCGGCGCCACGTCATCCGCTGCGCGAGGCTCAGAGGCCACTTCCGTGCGCGGTGG 471  
 Db 7230 TCGGCGCCACGTCATCCGCTGCGCGAGGCTCAGAGGCCACTTCCGTGCGCGGTGG 7171  
 QY 472 GTGGTCCCGGAGCGGCGAGTGGTCCAGAGGCGGCGGCTCGGTGAGCCGG 531  
 Db 7170 GCGGCTACCCGCAACCTGCTGTGAGCTGGTCCAAAGATGGCGGCGCTAGGTGACACAG 7111  
 QY 532 ACGGCCCCCGGCTGCGGTGAGGAGCTGCGCGAGGCAAGTGCCTCGCATTCGGGCGG 591  
 Db 7110 ATGCCCCCAGCTGCGGTGGAAGATCGCGAGAGGCGAGCGCTTCGATCCGTCGG 7051  
 QY 592 CGGCGCGCGGAGCGGCGGCACTTACGAGGTCCGCGCGGAGAACCCGCTGGGCGTGC 651  
 Db 7050 CAAGGCGCTCGGATGGTGGCACTACGAAAGTTCAGAGAGAACCCACTGGGCTCGGCCA 6991  
 QY 652 CGGCGCGCGGCGCTAGTGGTGGACTCGGACGCGCGGAGACAGCGCCAGCGCGCGGGA 711  
 Db 6990 GCGTGGCGCGCTCTGCTGTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 6934  
 QY 712 CTTCAACGCGCGGCTCTGCGGCACTGCGAGCGCGCGGCGAGGCTATCGCGCGGAGG 771  
 Db 6933 CTTCCGTTGCGACGCTCTGCGGCACTGCGAGCAGCGCGGCGAGGCTATCGCGCGAGG 6874  
 QY 772 GCGCGCGCGCTACCGCGGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 831  
 Db 6873 GCATACCTCTTCTCCACCTGCTGGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6814  
 QY 832 ACGCGCGCGCTCAGTGTCT 849  
 Db 6813 ACGCGCGCTCAGTGTCT 6796

RESULT 6  
 HSA314896 HSA314896 7531 bp DNA linear PRI 21-MAY-2002  
 LOCUS Homo sapiens partial OBSCN gene for obscurin, exons 1 and 2.  
 DEFINITION  
 ACCESSION AJ314896  
 VERSION AJ314896.1 GI:21104325  
 KEYWORDS OBSCN gene; obscurin.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
 Young, P., Ehler, E. and Gautel, M.  
 Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere assembly

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 7531)  
 AUTHORS Gautel, M.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-MAY-2001) Gautel M.S., Physikalische Biochemie,  
 Max-Planck-Institut fuer molekulare Physiologie, Otto-Hahn-Strasse  
 11, Dortmund, 44227, GERMANY

FEATURES  
 source  
 1..7531  
 Location/Qualifiers

gene  
 3420..7531  
 /gene="OBSCN"  
 mRNA Join(3420..3445,7022..7531)  
 /gene="OBSCN"  
 3420..3445  
 /gene="OBSCN"  
 /number=1  
 3420..3445  
 /gene="OBSCN"  
 3446..7021  
 /gene="OBSCN"  
 /number=1  
 intron

7022..7531  
 /gene="OBSCN"  
 /number=2  
 7044..>7531  
 /gene="OBSCN"  
 /function="titin and calmodulin binding Rho-GEF protein"  
 /codon\_start=1  
 /product="Obscurin"  
 /protein\_id="CAC85745.1"  
 /db\_xref="GI:21104326"  
 /db\_xref="SPTREMBL:Q8NH99"  
 /translation="MDQPFSGAPRFLTRPKAFVSVGKDATLSQIWNPTPOVSWE  
 XDCQPVAAAGAFRLAQDGLYRLTILDALGDSQYVCRARNAIGFAAVALQVDAE  
 AACAEAPHFLRLPTISRVRGSEATFCRVGSGSPRAVSWKDRRLGLGPDGPRV  
 EEL"

## ORIGIN

Query Match 47.2%; Score 505; DB 9; Length 7531;  
 Best Local Similarity 98.1%; Pred. No. 7,4e-52;  
 Matches 511; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
 QY 39 CCCCCATAGAGAGAGTCCCGCGCCACCGCTCATGTGATCAGCCACAGTTTCAGCGGGCG 98  
 Db 7011 CCCACCTATTTCAGAGTCCCGCGCCACCGCTCATGTGATCAGCCACAGTTTCAGCGGGCG 7070  
 QY 99 CCGCGCTTTCTACCGCGCCCAAGGCTTCGTGTGTCGGTGGGCAAGGACGCCACCTC 158  
 Db 7071 CCGCGCTTTCTACCGCGCCCAAGGCTTCGTGTGTCGGTGGGCAAGGACGCCACCTC 7130  
 QY 159 AGTCTCCAGATCGTGGTAAATCCACGCCACAGTGTGAGTGGGAGAGGACACAGCGCG 218  
 Db 7131 AGTCTCCAGATCGTGGTAAATCCACGCCACAGTGTGAGTGGGAGAGGACACAGCGCG 7190  
 QY 219 GTGACGCGCGCGCGCTTCGCTTGCGCCAGAGCGGACCTTACCGCTTCACTATC 278  
 Db 7191 GTGCGCGCGCGCGCTTCGCTTGCGCCAGAGCGGACCTTACCGCTTCACTATC 7250  
 QY 279 CTGACCTTGGCGCTGGGCGAGACAGTGGCAATACGTGTGCGCGCGGCAATGCGCATAGGC 338  
 Db 7251 CTGACCTTGGCGCTGGGCGAGTGGCAATACGTGTGCGCGCGGCAATGCGCATAGGC 7310  
 QY 339 GAGGCTTTCGTGCGTGGGCTTCGAGTGGACCGGAGCGCGCTGCGCGAGCGCG 398  
 Db 7311 GAGGCTTTCGTGCGTGGGCTTCGAGTGGACCGGAGCGCGCTGCGCGAGCGCG 7370  
 QY 399 CCGCACTTCTCTGCGCGCCACCGTCCATCCGCTGCGCGAGGCTCAGAGGCCACCTTC 458  
 Db 7371 CCGCACTTCTCTGCGCGCCACCGTCCATCCGCTGCGCGAGGCTCAGAGGCCACCTTC 7430  
 QY 459 CGTCCCGCGTGGTGGCTCCCGAGCGCGGACGTGAGTGTCCAGAGCGCGCGCG 518  
 Db 7431 CGTCCCGCGTGGTGGCTCCCGAGCGCGGACGTGAGTGTCCAGAGCGCGCGCG 7490  
 QY 519 CTGGGTGAGCCCGCGCGCGCGCGCGTGGCGTGGAGGAGCT 559  
 Db 7491 CTGGGTGAGCCCGCGCGCGCGCGCGTGGCGTGGAGGAGCT 7531

## RESULT 7

MMU409657

LOCUS

Mus musculus STS AGG4, sequence tagged site.

DEFINITION

ACCESSION

AJ409657

VERSION

AJ409657.1 GI:15131169

KEYWORDS

STS: repetitive DNA.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

1

Auchincloss, C.A.

Thesis (2001) Department of Medicine, University of Edinburgh,

Edinburgh, United Kingdom

```

REFERENCE 2
AUTHORS Auchincloss, C.A. and Abbott, C.M.
TITLE Isolation and mapping of novel mouse CpG island clones containing trinucleotide repeats, and demonstration of novel alleles in inbred strains
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1198)
AUTHORS Auchincloss, C.A.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-2001) Auchincloss C.A., Medical Genetics Section, Department of Medical sciences, Edinburgh University, molecular medicine centre, western general hospital, Edinburgh, EH4 2XU, UNITED KINGDOM

FEATURES             Location/Qualifiers
     source            1..1198
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="MF1/129"
                     /db_xref="taxon:10090"
                     /map="N/A"
                     /clones="39"
                     /sex="male"
                     /clone.lib="CpG island"
     STS               1..1198
                     /standard_name="AGG4"
     repeat_region     851..874
                     /evidence=experimental
                     /rpt_family="trinucleotide repeat"
                     /rpt_type=TANDEM

ORIGIN
Query Match          27.2%; Score 291.4; DB 11; Length 1198;
Best Local Similarity 72.0%; Pred. No. 5e-26;
Matches 503; Conservative 0; Mismatches 136; Indels 60; Gaps 7;

QY 52 AGGTCCCGCGCCACCGTCATGGATAGCCACAGTTTCAGCGGGGCGCCCGCTTTCTCA 111
DB 107 AGTCCCCCAACCCCACTCACATAATGGACCACTCTTCAGCGGACGACCCCGCTTCTGA 166
QY 112 CCGCGCCCAAGCGCTTCGTGTGTCGGTGGCAAGGACGCCACCTCAGTGCAGATCG 171
DB 167 CGCGGCCAAGGCTTT-GTGTATCTGTGTGAAGATGCCACCTGAGCTGCCAGATCG 225
QY 172 TGGGTAAATCCACGCGACAGTGTAGCTGGGAGAGGACCAAGCAGCCCGGTGACGCGCGCG 231
DB 226 TGGGCAACCCCAACGCGACACGTGAGCTGGGAGAGGACCGGACGACGACGAGGACGAG 285
QY 232 CGCGCTTCGCTTGGCCAGGACGCGACCTCTACCGCTCACTATCTTGACCTGGCG 291
DB 286 CGCGC-TCCGCTTGGCCAGGACGCGGATGTATACCGCTCACAAATCTCGATCTGGCG 344
QY 292 TGGGCGACAGTGGGCAATACGTGTGCCGCGCGCAATGCCATAGGCGAGCCCTTCGCTG 351
DB 345 TAGGTGACAGCGGCGAGTATGTGTGCGCGAGGACGCCATAGGCGAGGCGCTTGGC-G 403
QY 352 CGTGGGCTGAGGTGACGCGGAGGCGCGGTGGCGCGAGCGAGCGCGCGGCACTTCCTGC 411
DB 404 CGTGGGCTGAGGTGAGTGGAGGCGCGGAGGCGCGGAGGCGCGGAGGCGCGGAGGCGG 440
QY 412 TGGCGCCACGTCATCCGCTGCGCGAGGCTCAGAGGCGACCTTCGCTGCGCGCGTGG 471
DB 441 -----GAGCAGGACGCCACCTTCGACGTGCGGTG 471
QY 472 GTGGCTCCCGAGGCGCGAGTGTGTCGTCAGAGGACGCGCGCGCTGGGTGAGCCCG 531
DB 472 GAGGCTCAACCAACTGTGTGAGTGTGTCGTCAGAGGATGGGCGCG-CTGGGCGCCACCG 530
QY 532 ACGGCGCCCGGTGCGCTGAGGAGCTCGCGAGGAGTGCCTGCGCATTCGGCGG 591
DB 531 ATGCCCCCATGTGCTGTGAAGAACACGGGAGTCGAGCGCGCTTCGCATCCGGTCGG 590
QY 592 CGCGCGCGCGACGCGCGCACTTACGAGGTCCGCGCGAGAACCCCGCTGGCGCGCTGCCA 651

```

```

DB 591 CAAGGCTCGCATGGTGGCACCTACGAGTCCGAGCAGAAATCGGCTGGCTCTG-CA 649
QY 652 GCGCGCGCGCGCGCTAGTGTGAGTCTGGACCGCGGACACGCGCCAGCGCGCGCGGA 711
DB 650 GCGCTGCCGCGCTCTGGTTGTGACTCGGACGCGGAGG---TTGCCGACACCGCGGCA 706
QY 712 CQTCACCGCGCGCGCTCTGGGCGCACCTGCGACGCGCGG 750
DB 707 CQTCACCGCGCGCGCTCTGGGCGCACCTGCGACGCGG 745

RESULT 8
AX714317/c
LOCUS AX714317 2884 bp DNA linear PAT 15-APR-2003
DEFINITION Sequence 1001 from Patent EP1293569.
ACCESSION AX714317
VERSION AX714317.1 GI:29889269
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
MAMMALIA; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1
AUTHORS Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuho, Y.
TITLE Full-length cDNAs
JOURNAL Patent: EP 1293569-A 1001 19-MAR-2003;
Helix Research Institute (JP); Research Association for Biotechnology (JP)
FEATURES             Location/Qualifiers
     source            1..2884
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"

ORIGIN
Query Match          23.8%; Score 255; DB 6; Length 2884;
Best Local Similarity 100.0%; Pred. No. 8.4e-22;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 805 GCACCTGCGAGTGAAGGACGCGGCTCAGCTGCTACGTGACCGGCGAGC 864
DB 634 GCACCTGCGAGTGAAGGACGCGGCTCAGCTGCTACGTGACCGGCGAGC 575
QY 865 CCAAGCCGAGCGGTGTGGAAGAGCGCGGCTGTGACCGAGGCGCGCGCAGC 924
DB 574 CCAAGCCGAGCGGTGTGGAAGAGCGCGGCTGTGACCGAGGCGCGCGCAGC 515
QY 925 TGTGTACGAGGACGCGGAGGAACTTCGTGCTCAAGATCCTCTTCTCAAGCAGTCGG 984
DB 514 TGTGTACGAGGACGCGGAGGAACTTCGTGCTCAAGATCCTCTTCTCAAGCAGTCGG 455
QY 985 ACCGCGGCTCTACACCTGCACGCGTCCAACTCTGTGGGCGAGCCTACAGCTCTGTGC 1044
DB 454 ACCGCGGCTCTACACCTGCACGCGTCCAACTCTGTGGGCGAGCTACAGCTCTGTGC 395
QY 1045 TGTGTGTAGTGGCG 1059
DB 394 TGTGTGTAGTGGCG 380

RESULT 9
AK056556/c
LOCUS AK056556 2884 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ31994 fis, clone NT2RP7009215.
ACCESSION AK056556
VERSION AK056556.1 GI:16551990
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
EUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

RESULT 10	BC007201	3316 bp	mRNA	linear	PRI 12-JUL-2001
LOCUS	BC007201	3316 bp	mRNA	linear	PRI 12-JUL-2001
DEFINITION	BC007201	3316 bp	mRNA	linear	PRI 12-JUL-2001
ACCESSION	BC007201	3316 bp	mRNA	linear	PRI 12-JUL-2001
VERSION	BC007201.1	GI:13938169			
KEYWORDS	BC007201.1	GI:13938169			
SOURCE	BC007201.1	GI:13938169			
ORGANISM	BC007201.1	GI:13938169			
REFERENCE	BC007201.1	GI:13938169			
AUTHORS	BC007201.1	GI:13938169			
TITLE	BC007201.1	GI:13938169			
JOURNAL	BC007201.1	GI:13938169			
REMARK	BC007201.1	GI:13938169			
COMMENT	BC007201.1	GI:13938169			
FEATURES	BC007201.1	GI:13938169			
source	BC007201.1	GI:13938169			
CDS	BC007201.1	GI:13938169			

ORIGIN		WSRPPWRTAGTE"	
Query Match		18.6%; Score 199.2; DB 9; Length 3316;	
Best Local Similarity		51.9%; Pred. No. 3.8e-15;	
Matches		556; Conservative 0; Mismatches 498; Indels 18; Gaps 4;	
QY	11	CAGCCCACTCCGCGCTGCCAGAGCCCAATAGAGAGAGTCCCGCCCGCCACCGT	70
DB	5	CCGACACCCCTGTCGCCCCCAACCGTTCGCGCGTGCAGGCGGTGCGCCGCCCGGATGAA	64
QY	71	CATGGATCAGCACAGTTTCAGCGGGGGCCCGCTTTCTCACCGCGGCCCAAGCGCTTCGT	130
DB	65	GCGAGCTCGGGGATCAGGGAGAGCCCGCTTCCTCGCGTTCGCGGGCTGTGGC	124
QY	131	GGTGTGGTGGGAGAGAGCCACCTTCAGTTCACAGATGTGGTAAATCCACGCCACA	190
DB	125	GGTGTAAAGTGGCCCGAGGCGGAGCTCAAGTGGTGTCTGCGGGAGCGCGCGCTGT	184
QY	191	GTCAGCTGGAGAGACCAAGCGGTGACGGCGCGCGCTTCGCTCTGGGCCA	250
DB	185	AGTGTGGAGAGGGGGGAGAGCTGGCGGCTTCGAAAGCGCTGAGCTTCGCGC	244
QY	251	GGAGCGGACCTTACCGCTCTATCTGGACCTGGCGCTGGGCGACAGTGGGCAATA	310
DB	245	GGAGCGGCGGAGCAGCGCTCTGTCAGCCGCGCACTGCCACCGAGCGGGGTCTA	304
QY	311	CGTGTGCGCGCGGCAATGATAGCGAGCGCTTCGTCGCGTGGCGCTTCGACAGTGA	370
DB	305	CGTGTGCGCGCGGCAACGCGCGGCGGAGGCTTACGCGCGCGCGCTCACCGTCT	364
QY	371	CGCGAGCGCGCTGCGCGGAGCGCGCACTTCTGTGCGGCGCCACGTCATCCG	430
DB	365	GGAGCGCGCGCTCCGACCCGAGTGCAGCGCGCGGAGCGCGCTGCATCGCGGG	424
QY	431	CGTGGCGAGGCTCAGAGGCACTTCCGTGCGG-----CGTGGTGGCTCCCGAG	484
DB	425	GTCGCGGAGGCGCGCGGCTTCTTCAAGGGGCGCTCGATCCAGTGGGTGTGCGGG	484
QY	485	CGCGGAGTGTGCTGAGTGTGCGGCGCGCTGGTGAGCGCGCGCGCGCGCGCGT	544
DB	485	GGCGAGTGTGCTGAGTGTGCGGCGCGGCGCTCCCGAGGCCACACTGACTGGA	544
QY	545	GCGGTGAGAGTGTGCGGAGGCAAGTGCCTGCGGATTCGGGCGCGCGCGCGCGGA	604
DB	545	GAAAGACGGATGGCGCTGAGCAAGTGTGGGACAGCAGCACTTCGCGCTCCAGCGCGG	604
QY	605	CGCGGCACTTACAGGTTCGCGCGGAGAACCGCTGGC--GCTGCAGCGCGCGCGCG	662
DB	605	CGCGCGGAGGAGCGCGCGCGCGAGCTGGCATTCGCGATCTGGCGGCTCGCTGCC	664
QY	663	GCGCTAGTG-GTGGACTCGGCGCGGACACGCGCCAGCGCGCGCGCGGACCTCCAGCG	721
DB	665	GGATTCCGCGCTACGTGTGCGACCGCGCAAGCGCGCACCGCGCGCGCGCGCGG	724
QY	722	CGCGCTCTGGGCACTTGCAGCGCGCGGCGAGGCTATCGCGCGGAGCGCGCGCGCG	781
DB	725	GCTGTCCAGGTGACCAAGCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCG	784
QY	782	CTC-----ACCGCCAGCACCGGACCGCGCACTTCGACGCTGACTGAAGGCAAGCA	832
DB	785	GCGGTGTGGAGCGCTCAAGTGGCGCTTAAGACCTTCTGGGTGAACGAGGCAAGCA	844
QY	833	CGCGCGCTCAGCTGCTAGTACCGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCG	892
DB	845	CGCCAAAGTTCGCTCTAGTGTGGCAAGCGCGCGCGCGCGCGCGCGCGCGCG	904
QY	893	CGCGCGCTGTGTACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	952
DB	905	GGCGCGCGCGCTGCTCCCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG	964
QY	953	CGTGTCAAGATCTCTTGTCAAGCAGTGGAGCGCGCGCGCTTACCTCAGCGCGCTC	1012

DB	965	CGTGTCAAGGTGCTCTACTCCAGGCAAGGATCGTGGGTCTACGTCTGCGCGCGCG	1024
QY	1013	CAACTCTGTGGCCAGACCTACAGCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT	1064
DB	1025	CAACTCGCGCGCGCCAGACGCTCAGTGCCTGACGTGACGTGAAAGAGCGCC	1076
RESULT 11			
AC009955/c			
LOCUS			
Homo sapiens BAC clone RP11-256I23 from 2, complete sequence.			
AC009955			
VERSION			
AC009955.4 GI:9581957			
KEYWORDS			
HTG.			
SOURCE			
Homo sapiens (human)			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE			
1 (bases 1 to 135033)			
Sulston, J.B. and Waterston, R.			
Toward a complete human genome sequence			
Genome Res. 8 (11), 1097-1108 (1998)			
MEDLINE			
99063792			
PUBMED			
9847074			
REFERENCE			
2 (bases 1 to 135033)			
Du, H., Maupin, R., Hawkins, M. and Hodges, J.			
The sequence of Homo sapiens BAC clone RP11-256I23			
Unpublished			
REFERENCE			
3 (bases 1 to 135033)			
Waterston, R.H.			
Direct Submission			
TITLE			
Submitted (08-SEP-1999) Genome Sequencing Center, Washington			
University School of Medicine, 4444 Forest Park Parkway, St. Louis,			
MO 63108, USA			
REFERENCE			
4 (bases 1 to 135033)			
Waterston, R.H.			
Direct Submission			
TITLE			
Submitted (29-JUL-2000) Genome Sequencing Center, Washington			
University School of Medicine, 4444 Forest Park Parkway, St. Louis,			
MO 63108, USA			
REFERENCE			
5 (bases 1 to 135033)			
Waterston, R.H.			
Direct Submission			
TITLE			
Submitted (10-SEP-2000) Genome Sequencing Center, Washington			
University School of Medicine, 4444 Forest Park Parkway, St. Louis,			
MO 63108, USA			
REFERENCE			
6 (bases 1 to 135033)			
Waterston, R.			
Direct Submission			
TITLE			
Submitted (14-OCT-2000) Department of Genetics, Washington			
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA			
REFERENCE			
7 (bases 1 to 135033)			
Waterston, R.			
Direct Submission			
TITLE			
Submitted (08-NOV-2000) Department of Genetics, Washington			
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA			
On Jul 29, 2000 this sequence version replaced gi:8569874.			
COMMENT			
----- Genome Center			
Center: Washington University Genome Sequencing Center			
Center code: WUGSC			
Web site: http://genome.wustl.edu/gsc			
Contact: sapiens@wustl.edu			
----- Summary Statistics			
-----			
Center project name: H_NH0256123			
-----			

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/qsc>

**SOURCE INFORMATION:**

For construction of the BAC library was made from the blood of one male donor as described by Casagawa, K., and Y. Tsubota, B. Resgen, E., Tateno, M., Casagawa, J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-51207, 200 base pair overlap. Actual start of this clone is at base position 1 of RP11-256123; actual end is at base position 52326 of RP11-51207.

## FEATURES

[illegible]

Db 42125 CGCGATGAAGCGAGCTCGGGGATCAGGGAGCCCCCGCTGCTTCGCGCTCCCGCG 42066  
 Qy 122 GGCCTTCGTGTCGTGGGCAAGGACGCCACCTCAGTGCAGATCGTGGGTATCC 181  
 Db 42065 GCCTGTGCGGTGTAAAGTGGCGCGGAGGCGGAGCTCAAGTGTGCTGCTGGGAGCC 42006  
 Qy 182 CACGCCACAGTGTAGCTGGGAGAGGACACGAGCGGTGACGCGCGCGCGCTTCGG 241  
 Db 42005 GCCGCTGTAGTGTGGGAGAGGCGGCGAGCAGCTGGCGCTCGGAACGCTGAG 41946  
 Qy 242 TCTGGCCAGGACGGCGACCTTACCGCTCCTACTATCTGAGACTGCGCTGGGCGACAG 301  
 Db 41945 CTTCCCGCGGACCGCGGAGCAGCGCTCTGCTGACCGCGGCACTGCCACCGACGC 41886  
 Qy 302 TGGCAATACGTGTGCGCGCGCAATGCAATAGGCGAGCGCTTCTGCTCGCGTGGCGCT 361  
 Db 41885 GGGGTCTACGTGTGCGCGCGCGACGCGGCGGAGCGCTACGCGGCGCGCGCT 41825  
 Qy 362 GCAGGTGAGCGCGAGCGCGGTGCGCGGAGCGCGGACTTCTGCTGCGGCGCCAC 421  
 Db 41825 CACGCTGTGAGCGCGCGCTCGGACCGGAGCTGACGCGCGCGGCGCGCTGCC 41766  
 Qy 422 GTCCATCGGTGCGCGAGGCTCAGAGGCGACCTTCGCTGCGTGG 475  
 Db 41765 ATCCCGGGTCCGGGAGGCGCGCGCTTCTTCTCAGGGGCTCGATCCAGTGGT 41706  
 Qy 476 CTCGCCGAGCGCGGAGTGTGCTCCAGGACGCGCGCGCTGCGTGGTGGCGCGACGG 535  
 Db 41705 GCTCGGGGCGGAGTGTGCTGACGTGCGGCGCGGCGCGCTCCCGGAGCGCACACT 41646  
 Qy 536 CCCCGGTGCGGTGAGGAGCTCGGCGAGGCAAGTGCCTGCGCATTCGGGCGCGCG 595  
 Db 41645 GTACTGGGAGGACCGGATGCGGCTGGAGCAAGTGTGGACAGGACCACTTCGCGCT 41586  
 Qy 596 CGCGCGGACCGCGGCACTTACGAGTCCGCGCGGAGAACCGCTGGGC--GTTGCCAGC 653  
 Db 41585 CCAGCGGCGCGCGGAGGACGCGCGCGCGAGCTGCGCATCTGGCGG 41526  
 Qy 654 GCGCGCG--CGCGGTAGTGTGAGTCTGAGCGCGCGGACGCGCGCGCGCGGAC 712  
 Db 41525 TCGGCTGCGGATTCGCGGCTTACGTGTGCGACGCGCGCAACGCGCGCGCGCGCA 41466  
 Qy 713 CTCACGCGCGCTCTCGGCGCACCTGACGCGCGCGGAGGCTATGCGCGCGGAGG 772  
 Db 41465 GCGGGGCGGTGCTTCAGGTGACACGAGCCCCCGAGAGCCGCGCGGACCCCGACGA 41406  
 Qy 773 CGCCCCCGCTC-----ACCGCCGACCGCGGACCGCGCACCTGCGACGCTGACTGA 823  
 Db 41405 GCGCGCGCGCGGTGGAGCGGCTCAAGTGGCGGCTAAGACCTTCTGGGTGAACGA 41346  
 Qy 824 AGGCAAGCACGCGCGCTCAGTCTAGTACGCGGAGCGCGGAGCGCGCGGAGCGGTG 883  
 Db 41345 GGGCAAGCACCAAGTTCGCTGCTAGTGTGGGCAAGCGCGGAGCGCGGAGTCTGAATG 41286  
 Qy 884 GAAGAAGGACCGCGGAGTGTGACCGGCGCGGCGGCGCACGCTGTGTGTACGAGGACGCGCA 943  
 Db 41285 GCACGTGGAGGCGCGCGCTGCTCCGAGCGCGCGCGCTCTATGTACGCGGACCGCA 41226  
 Qy 944 GGAGAACTGCTGCTCAAGATCTCTTCTGCAAGCAGTTCGACCGCGCGCTTACACTG 1003  
 Db 41225 CGCGGCTTCTGCTCAAGTGTCTTACTGCGCGGCAAGGATCGTGGCTCTACGCTG 41166  
 Qy 1004 CACGCGTCCAACTCTGCGGCGGAGACCTACAGCTCTGTGCTGTGTGCTGCGCGGAGCC 1053  
 Db 41165 GCGCGCGGCAACTCGCGGCGGCGGAGCGCTCAGTCCGCGTGCAGCTGCGGAGGATC 41106  
 Qy 1064 GCGCG 1068  
 Db 41105 GCGCG 41101

RESULT 12

AC040991  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

TITLE  
 JOURNAL  
 COMMENT

AC040991 186883 bp DNA linear HTG 24-AUG-2002  
 Homo sapiens chromosome 2 clone RP11-123E11 map 2, WORKING DRAFT  
 SEQUENCE, 17 unordered pieces.  
 AC040991  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 AC040991.2 GI:8014564  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 186883)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 2, clone RP11-123E11  
 Unpublished  
 2 (bases 1 to 186883)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Balwin,J., Barna,N., Bastien,V., Bada,F.,  
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,  
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczkv,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Melrhy,T., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisaní,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 186883)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Balwin,J., Barna,N., Bastien,V., Bada,F.,  
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,  
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczkv,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Melrhy,T., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisaní,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On May 22, 2000 this sequence version replaced gi:7534193.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Web site: http://www-seq.wi.mit.edu

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L9492

Center clone name: 123 E 11

## ----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: phrap; version 0.960731

Consensus quality: 175185 bases at least Q40

Consensus quality:	181490	bases at least	Q30
Consensus quality:	173100	bases at least	Q30

Consensus quality:	183848	passes at least	Q20
Consensus quality:	101150	passes at least	Q30

Insert size: 190000; agarose-fp

**Insert size:** 185283; sum-of-squares: 1000; aggregate-p

Insert size: 1022bp, sum-OI-concigs  
Quality coverage: 4.3 in Q20 bases: agarose-fro

Quality coverage: 4.3 in Q20 bases; agarose-lp  
Quality coverage: 4.4 in Q20 bases; sum-of-contigs

----- Quality Control

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

* 1	1462:	contig of 1462 bp in length
* 1463	1562:	gap of 100 bp
* 1563	4486:	contig of 2324 bp in length
* 4487	4586:	gap of 100 bp
* 4587	9251:	contig of 4665 bp in length
* 9252	9351:	gap of 100 bp
* 9352	12768:	contig of 3417 bp in length
* 12769	13868:	gap of 100 bp
* 12859	17380:	contig of 4512 bp in length
* 17381	17480:	gap of 100 bp
* 17481	22233:	contig of 4753 bp in length
* 22234	22333:	gap of 100 bp
* 22334	25600:	contig of 3267 bp in length
* 25601	25700:	gap of 100 bp
* 25701	31826:	contig of 6226 bp in length
* 31827	31926:	gap of 100 bp
* 31927	41356:	contig of 9430 bp in length
* 41357	41456:	gap of 100 bp
* 41457	53195:	contig of 41739 bp in length
* 53196	53295:	gap of 100 bp
* 53296	64559:	contig of 11264 bp in length
* 64560	64650:	gap of 100 bp
* 64650	79824:	contig of 15165 bp in length
* 79825	79924:	gap of 100 bp
* 79925	99903:	contig of 13985 bp in length
* 99910	100009:	gap of 100 bp
* 100010	121977:	contig of 21968 bp in length
* 121978	122077:	gap of 100 bp
* 122078	143182:	contig of 21105 bp in length
* 143183	143282:	gap of 100 bp
* 143283	164587:	contig of 21305 bp in length
* 164588	164687:	gap of 100 bp
* 164688	195883:	contig of 32196 bp in length

FEATURES  
SOURCE

```
1. 186883
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
```

	Human	Male	BAC
misc_feature	11462	assembly_fragment"	
misc_feature	1563	4486	
misc_feature	1563	"note="assembly_fragment"	
misc_feature	4587	9251	
misc_feature	9353	"note="assembly_fragment"	
misc_feature	9353	12768	

misc_feature	12869..17380	/note="assembly_fragment"
	clone_end:T7	
misc_feature	17481..22233	/note="assembly_fragment"
	vector_side:left	
misc_feature	22334..25600	/note="assembly_fragment"
	clone_end:SP6	
misc_feature	25701..31826	/note="assembly_fragment"
	vector_side:right	
misc_feature	31927..41356	/note="assembly_fragment"
	clone_end:SP6	
misc_feature	41457..53195	/note="assembly_fragment"
	clone_end:SP6	
misc_feature	53296..64559	/note="assembly_fragment"
	clone_end:SP6	
misc_feature	64660..79824	/note="assembly_fragment"
	clone_end:SP6	
misc_feature	79925..99909	/note="assembly_fragment"
	clone_end:SP6	
misc_feature	100010..121977	/note="assembly_fragment"
	clone_end:SP6	
misc_feature	122078..143182	/note="assembly_fragment"
	clone_end:SP6	
misc_feature	143283..164587	/note="assembly_fragment"
	clone_end:SP6	
misc_feature	164688..186883	/note="assembly_fragment"
	clone_end:SP6	
ORIGIN		
Query Match 18.5%; Score 197.8; DB 2; Length 186883;		
Best Local Similarity 51.6%; Pred. No. 1.4e-15;		
Matches 560; Conservative 0; Mismatches 507; Indels 18; Gaps 4;		
Qy	2	GCTTACAGAGCCACACTCGGCGCGTGCCAGAGCCGCCATAGAGAGTCCCGCG 61
Db	109463	GTCGCGCTTCGACACCCCGCTGCCCGCCACCGTTGCGCTGAGCGGTGCGCGCG 109522
Qy	62	CGCACCGTCATGGATCAGCCACAGTTTCAGCGGGCGCCCGCTTCTCACCGGCCCAA 121
Db	109523	CGCGATGAAGCGAGCTCGGGGATCAGGGAGCCCGCGTTCCTGCGCTTCCGCG 109582
Qy	122	GGCTTCGTGTTGTCGTGGGCAAGAGCGACCCCTCAGTGTGCGATCGTGGTAATCC 181
Db	109583	GCTGTGCGGGTGTAACTGCGCGCGAGCGCGAGCTCAAGTGGCTGTCTCGGGAGCG 109642
Qy	182	CACGCCACAGTGAGCTCGGGAGAGGACAGCAGCGGTGACGCGCGCGCGCTTCG 241
Db	109643	CGCGCTGTAGTGTGCGAGAGAGGGGCGGACAGCTGCGCGCTTCGAGACCTGAG 109702
Qy	242	TCTGCCAGGACGCGCACTTACCGCCTCACTATCCTTGAGACTGCGCTGGGCGACAG 301
Db	109703	CTTCCGCGCAGCGCGCGAGCAGCGCTGCTGACCGCGCACTGCCACCGACGC 109762
Qy	302	TGGGCAATACGTGTCCGCGCGCAATGATGATAGGCGAGCGCTTCGCTGGCGCT 361
Db	109763	GGGGTCTACGTGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 109822
Qy	362	GCAGGTGAGCGCGGAGGCGCGTGCAGAGCGCGCGCACTTCTCTCTCGGCGCCAC 421
Db	109823	CACCGTGTGAGCGCGCGCGCTCCGACCCCGAGCTGACGCCCGCGAGCGCGCTGC 109882
Qy	422	GTCCATCGCGTGCAGGAGCTCAGAGGCACTTCTCCGTCGCG-----CGTGGGTG 475
Db	109883	ATCGCGGGGTTCGGGAGAGGCGCGCGCGCTTCTCTCAGCGGCGCTCTGATCCAGTGGT 109942
Qy	476	CTCCCGAGGCGCGAGTGTGTCGTCGTCGCGCGCGCGCGCGCGCGCGCGCGCG 535
Db	109943	GCTCGCGGGGCGAGGTGTGTCGTCGTCGCGCGCGCGCGCGCGCGCGCGCGCG 110002





```

*****
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 224074: contig of 224074 bp in length
* 224075 224174: gap of unknown length
* 224175 225471: contig of 1297 bp in length.
FEATURES             source
     Location/Qualifiers
         1..225471
             /organism="Rattus norvegicus"
             /mol_type="genomic DNA"
             /db_xref="taxon:10116"
             /clone="CH230-61F12"
         1..1154
             /note="wgs contig"
         17776..18866
             /note="wgs contig"
         134828..136148
             /note="wgs contig"
         222128..224074
             /note="wgs contig"

misc_feature
misc_feature
misc_feature
misc_feature

ORIGIN
Query Match          16.9%; Score 181; DB 2; Length 225471;
Best Local Similarity 52.18; Pred. No. 1.4e-13;
Matches 509; Conservative 0; Mismatches 450; Indels 18; Gaps 4;

QY 97 CGCCCGGTTTCTACCGGCGCCAGGCTTGTGTGTCGGTGGGAGGAGGCGCACCC 156
Db 142585 CCCCCTGTTTCTACCGGCGCCAGGCTTGTGTGTCGGTGGGAGGAGGCGCACCC 156

QY 157 TCAGCTGCAGATCGTGGGTAATCCACGCGCACAGGTGAGCTGGGAGAGGACGACGAGC 216
Db 142645 TCATATCGTGGTCTCTGGAGAGCGCGCGCCACCGCTCTTGTGGAGAGAGGCGGCGAGC 142704

QY 217 CGGTAGCGCGCGCGCGGTCTCGTGTGCGCCAGGACGCGACCTCTACCGCTCACTA 276
Db 142705 AGCTGTGGTCTCGAGAGCGCTGAGCTTCCAGAGGACGCGCGCGCGCTGCTGT 142764

QY 277 TCCTGGACCTTGGCGCTGGGCGACAGTGGGCAATACGTGTGCGCGCGCGCAATGCCATAG 336
Db 142765 TGAGCGGCGGTCTCCACCGACGCGCGGCTTACGTGTGCGCGCGCGCGCGCGCGCG 142824

QY 337 GCGAGGCGCTTC-----GCTCCGCTGGGCTGTGAGTGGACGCGGAGGCGCGTGGCGCG 390
Db 142825 GAGAGGCGCTACGCGCGCGCGCGCTACCGCTGTGGAACCCCGCGCGCGCGCGAGC 142884

QY 391 AGCAGGCGCGCTTCTCTCTGCGGCCACGCTCCATCCGCGTGGCGGAGGCTCAGAGG 450
Db 142885 CCGAGTCTTCGAGTGTCCCGCGCGCGCTCCCGGACCGCGGAGGCGCGCGCGCTTTTC 142944

QY 451 CCACCTTCCGCTGCGCGTGGGCTGCTCCCGAGGCGCGGCGAGTGGTGTCCAGGAGC 510
Db 142945 TGACGGGACCCAGTCCAGTCCAGTGGTGTGCGAGGCGGAGGTTGGTGTGAGTGCACG 143004

QY 511 GCGGCGCGCTGGTGTGAGCCGACGCGCGCGCGCTGCGCGTGGAGGAGCTCGCGGAGGCAA 570
Db 143005 TGGAGGCGCTCCCGGAGCGCCAAAGCTGTGAGAGAGGAGTGGGCTTGGACGAG 143064

QY 571 GTGCGCTGCGCATTCGGGCGCGCGCGCGCGCGCGCGCACTTACGAGGTTCCGCGCG 630
Db 143065 TGTGGGACAGAGCGCACTACACGCTGGAGCGGACCGCGCGCGAGTGTGAGCGGCGCAA 143124

QY 631 AGAACCGCGTGGG--GCTGCGAGCGCGCGCG--CGGCGGTAGTGTGGACTGTGAGCGCG 687
Db 143125 GCCTGACGTTACGATCTCTGGCAGCGCGCGGTGCGCGGATTCGCGGGGTGTACGTGTGCCAGC 143184

```

```

QY 688 CGGACACGCGCAGCGCGCGCGCGCGCGCGCTCCAGCGCGCGCGCTCTCTGGCGCACCTGACGCGC 747
Db 143185 CCGCAACGCGCAGCGCGCGCGCGCGCGCGCTCTCCAGGTGCGCGCGCGCGCGCGCGCGCGC 143244

QY 748 GCGCGAGGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 798
Db 143245 AGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143304

QY 799 GCACGCGCAGCTGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 858
Db 143305 CGCCCAAGACCTTCTGGGTGAACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143364

QY 859 GCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 918
Db 143365 GCAAGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143424

QY 919 GCGAGCTGCTGCTACGAGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 978
Db 143425 CCGCGCTCATGTACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143484

QY 979 AGTCGACCGCGCGCGCTCTACACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1038
Db 143485 CCAAGGACCGTGGCGCTCTACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 143544

QY 1039 CTGTGCTGCTGTAGTG 1055
Db 143545 CGTTGAGCTGACGTG 143561

```

## RESULT 14

AC121633/c

LOCUS

DEFINITION

Rattus norvegicus clone CH230-62C23, WORKING DRAFT SEQUENCE, 6

unordered pieces.

ACCESSION

AC121633

VERSION

AC121633.4 GI:25137860

KEYWORDS

HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 266167)

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Georgopoulos, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunnarson, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, C.,

Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,

Hollins, B., Howells, S., Hui, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpatsky, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kovis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorenshew, L., Loulseg, H., Lozano, R. J., Lu, X., Ma, J.,

Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, S.,

Mawhinney, S., McLeod, M.P., McNeill, T.Z., Neenen, E.,

Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackeleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poinexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Staimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Direct Submission  
Unpublished  
2 (bases 1 to 266167)

Worley, K.C.

Submitted (21-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 266167)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:22856334.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly ('a' contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
-----  
Center project name: GWAS  
Center clone name: CH230-62C23  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 252717 bases at least Q40  
Consensus quality: 255419 bases at least Q30  
Consensus quality: 257320 bases at least Q20  
Estimated insert size: 258645; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will

\* be preserved.  
\* 1 256774: contig of 256774 bp in length  
\* 256775 256874: gap of unknown length  
\* 256875 258093: contig of 1219 bp in length  
\* 258094 258193: gap of unknown length  
\* 258194 259332: contig of 1339 bp in length  
\* 259333 259632: gap of unknown length  
\* 259633 261419: contig of 1787 bp in length  
\* 261420 262537: contig of 1018 bp in length  
\* 262538 262637: gap of unknown length  
\* 262638 266167: contig of 3530 bp in length.

#### FEATURES

source  
1..266167  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-62C23"

#### ORIGIN

Query Match 16.9%; Score 181; DB 2; Length 266167;  
Best Local Similarity 52.1%; Pred. No. 1.3e-13; Indels 18; Gaps 4;  
Matches 509; Conservative 0; Mismatches 450;  
QY 97 CGCCCCGCTTTCTCACCCCGGCCAAAGCCCTTCGTGTGTCTCGGTGGGCAAGGACGCCACCC 156  
Db 244163 CCCCCGTTTCTACGCTTCCCGCGCGCGTGTAGTGGCGGAGGCCGAGC 244104  
QY 157 TCAGCTGCCAGATCGTGGGTAAATCCACGCCACAGTGTAGCTGGAGAGAGGACCGACGC 216  
Db 244103 TCAAAATGCGTGTCTCTGGGAGAGCGCCGCCACCGCTCTTGTGGGAGAAAGCGCGCAGC 244044  
QY 217 CGGTGACGCGCGCGCGCTTCCGTCCTGGCCAGGACGGCGACCTCTACCCGCTCACTA 276  
Db 244043 AGCTGTGCGCTTCGGAGCGCCTGAGCTTTCAGAGAGCGCGCCCGACGAGCTGCTGT 243984  
QY 277 TCTTGACCTGCGCTGGCGCGACAGTGGGCAATACGTGTGCCCGCGCGCGCAATGCGCATAG 336  
Db 243983 TGAGCGCGCGTGTGCCACCGCGCGGTCTACGTGTCCGCGCGCGCAACCGCGCGCG 243924  
QY 337 GCGAGCGCTTC-----GCTGCGGTGGGCTCAGTGTGACCGGAGCGCGGTGCGCGCG 390  
Db 243923 GAGAGCGCTTACGCGCGCGCGCGTCACTGCTGAAACCCCGCGCGCGCGCGCGCGCGCG 243864  
QY 391 AGCAGCGCGCGCACTTCTGCTGCGGCGCCACGCTCCATCCGCTGCGCGAGGCTCAGAGG 450  
Db 243863 CCCAGTCTTCGAGTGTCCGCGCGCGCTCCGCGCACCGGGAGGCGCGCGCGGTGTTTC 243804  
QY 451 CCACCTTCGCTGCCCGGTGGGTGCTCCCGAGGCGCGCAGTGTGTCAGGAGCG 510  
Db 243803 TGACGGGACCCAGTCCAGTGGGTGCTCAGAGGGCGGAGGTGTGTGCTGCCAGG 243744  
QY 511 GCGCGCGCTGGGTGAGCCCGCGCGCGCGTGGCGGTGGAGAGCTCGCGAGGCGAA 570  
Db 243743 TGGGAGCGCTCCCGGAGCCCAAGCTGTACTGGGAGAGGATGGATGGCTTTGGACGAAG 243684  
QY 571 GTGCGCTGCGCAATTCGGGCG 630  
Db 243683 TGTGGGACAGCAGCCACTACAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 243624  
QY 631 AGAACCGCTGGGC--GCTGCCAGCGCGCGCG--CGGCGCTAGTGTGGGCTCGGAGCGCG 687  
Db 243623 GCTGACGTTTACGATCTCTGGCAGCGCGCGCTGCCCGATTCGGGGGTGTACGTGTGCACG 243564  
QY 688 CGGACACGGCG 747  
Db 243563 CCGCGAACCGCAGCGCCACCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 243504  
QY 748 GCGCGCAGGCTATGCGCGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 798  
Db 243503 AGAACCG 243444  
QY 799 GCACGCGCACCTGCACGCTGACTGAAGGCAACGACGCGCGCGCGCTCAGCTGTACGTGACCG 858



```
QY 277 TCCTGCACTGGCGTGGCGACAGTGGCAATACGTGTGCGCGCGCGCAATGCCATAG 336
Db 78241 TGAGCGGCGGCTGCCACCGCGGGGTCTACGTGTGCGCGCGCGCAACGCGGCG 78300
QY 337 GCGAGGCGCTTC-----GCTGCGGTGGGCTTCAGTGGAGCGGAGGCGCGTGCGCCG 390
Db 78301 GAGAGGCTACGCGGCGCGCGGTACCGTGTGGAACCGCGCGCGCGCGCGGAGC 78360
QY 391 AGCAGGCGCGCACTTCCTGCTGGGCGCCAGTCCATCCGCGTGGCGAGGGCTCAGAG 450
Db 78361 CCGAGTCTCCGAGTGTCCGTGCCACACCGGCGCACCGGGAGGGCGCGCCGAAATTCC 78420
QY 451 CCACCTTCGCTGCCGCGGTGGTGTCTCCCGAGCGCGGCGAGTGTGCTGCCAAGGACG 510
Db 78421 TGACGGGCGCCAGTCCAGTGGGTGTGCGAGGGGAGAGTGGTGTGACGTGCCAG 78480
QY 511 GCGCGCGCTGGGTAGCCCGACGGCCCCCGCGTGGCGGTGGAGAGCTGGCGAGGCAA 570
Db 78481 TGGGAGGCGCTTCGCGAGCGCCAAAGTGTACTGGGAGAGGATGGGATGGCTTGGACGAAG 78540
QY 571 GTGCGCTCGCATTCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 630
Db 78541 TATGGACAGACGACCATTCAGCTGGAGCCCGCGCGCGCGCGCGCGCGCGCGCG 78600
QY 631 AGAACCCGCTGGGCGCTCCAGCGCGCGCGCGCGCGCTA----GTGGTGGACTCGGACGCC 686
Db 78601 GCCTGACGTGGCG-ATCCTGGCGCACCGGTGCCCGATTCCGGGGTGTACGTGTCTAC 78659
QY 687 GCGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 746
Db 78660 GCCCGCAACGGCGACGGCGCACGGCGAGCGGCGCGCGCGCGCGCGCGCGCGCGCG 78719
QY 747 CGGCGCGAGGCTATGCGCGCGAGGCGCGCGCGCGCGCTC-----ACGCGCGCGCAC 797
Db 78720 GAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 78779
QY 798 GGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 857
Db 78780 GCGCGCAAGACCTTCTGGGTGAACGAGGCGCAAGCACGCGCAATTCGCGTGTCTAGTGATG 78839
QY 858 GCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 917
Db 78840 GGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 78899
QY 918 CGCCACGCGGTGTACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 977
Db 78900 CCGCGCGCTCATGTACCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 78959
QY 978 CAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1037
Db 78960 GCCAAGGACCGTGGGCTCTAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 79019
QY 1038 TCTGTGCTGCTGTAGTG 1055
Db 79020 GCGGTCCAGCTGCAGTG 79037
```

Search completed: September 19, 2004, 19:26:59  
Job time : 4372.03 secs

**This Page Blank (uspto)**